TEXAS A&M UNIVERSITY
PROJECT THEMIS

Technical Report #58

Statistical PERT: Improvements in the Determination of the Project Completion Time Distribution

by

Thomas C. Baker, Jr. and Robert L. Sielken Jr.

Texas A&M University
Office of Naval Research
Contract N00014-78-C-0426
Project NR047-179

Attachment I

Statistical PERT: Improvements in the Determination of the Project Completion Time Distribution

by

Thomas C. Baker, Jr. and Robert L. Sielken Jr.

THEMIS OPTIMIZATION RESEARCH PROGRAM
Technical Report No. 58
August 1978

INSTITUTE OF STATISTICS
Texas A&M University

Research conducted through the
Texas A&M Research Foundation
and sponsored by the
Office of Naval Research
Contract N00014-78-C-0426
Project NR047-179

ATTACHMENT II

Statistical PERT: Improvements in the Determination of the Project Completion Time Distribution

Thomas C. Baker, Jr. and Robert L. Sielken Jr.

ABSTRACT

This report develops improvements to a new project scheduling procedure, Statistical PERT, being developed at the Institute of Statistics, Texas A&M University. The project scheduling algorithm is a five step iterative procedure capable of determining a minimum cost project schedule when the activities making up the project have durations which are random variables. The cost of an activity is assumed to be a convex piecewise linear function of the activity's mean duration. The problem is to determine the activity mean durations which both minimize the total project cost and insure that the mean (or some specified percentile) of the corresponding project completion time distribution is less than or equal to a specified project deadline. The entire distribution of the project's completion time under the minimum cost schedule is a valuable by-product.

A critical step, Subnetwork Analysis, in the proposed procedure is improved and extended. Subnetwork Analysis determines an estimate of the duration distribution, $F(t)$, for each subnetwork identified in the previous steps. This estimate is extended to include an extrapolation of upper and lower bounds on $F(t)$. This report also develops a new sampling procedure which results in improved estimators for the bounds on $F(t)$.

TABLE OF CONTENTS

Section		Page
1.	A STATISTICAL APPROACH TO PROJECT SCHEDULING	1
	1.1 Introduction	1
	1.2 The Project Scheduling Problem	1
	1.3 Outline of the New Approach to Project Scheduling	4
	1.4 An Integrated System of Computer Programs	6
	1.5 The Determination of the Subnetwork Duration Distribution	7
2.	ANALYSIS OF A SUBNETWORK	9
	2.1 Introduction	9
	2.2 Formation of Clusters	11
	2.3 Bounding the Discrete Subnetwork Duration Distribution F	19
	2.3.1 Upper Bounds on F	19
	2.3.2 Lower Bounds on F	22
	2.3.3 The Tightness of the Bounds on F	23
	2.4 Using Sampling to Estimate the Upper and Lower Bounds on F	24
	2.5 Estimating F by Extrapolating Between the Upper and Lower Bounds on F	25
	2.6 A Summary of the Subnetwork Analysis Procedure	26
3.	SAMPLE-BASED ESTIMATORS FOR A DISCRETE DISTRIBUTION FUNCTION	29
	3.1 Introduction	29
	3.2 Some Proposed Estimators	30
	3.2.1 The Empirical Distribution Function, $G_1(t)$	33
	3.2.2 The Modified Empirical Distribution Function, $G_2(t)$	33

TABLE OF CONTENTS (continued)

Section		Page
	3.2.3 The Continuous Estimator, $G_3(t)$	33
	3.2.4 The Mixed Estimator, $G_4(t)$	34
	3.2.5 The Mixed Estimator, $G_5(t)$	34
	3.3 Criteria for a Good Estimator	34
	3.4 Choosing Between Simple Random or Systematic Sampling	36
	3.4.1 An Ordering Scheme	36
	3.4.2 Implementing the Ordering Scheme	41
	3.5 The Simulation Study	42
	3.5.1 A Comparison of $G_1(t)$, ..., $G_5(t)$ Under Systematic Sampling	45
	3.5.2 The Performance of $G_2(t)$ Under Systematic and Random Sampling	51
4.	ESTIMATION OF A DISCRETE DISTRIBUTION FUNCTION BY EXTRAPOLATING UPPER AND LOWER BOUNDS	53
	4.1 Introduction	53
	4.2 The Extrapolation Problem	53
	4.3 A Linear Programming Solution to the Extrapolation Problem	55
	4.4 An Example of the Linear Programming Solution	58
5.	POTENTIAL MODIFICATIONS OF THE SUBNETWORK ANALYSIS PROCEDURE	61
	5.1 Introduction	61
	5.2 Explicit Evaluation of the Subnetwork Duration Distribution	63
	5.3 Approximating the Subnetwork Duration Distribution $F(t)$	66

TABLE OF CONTENTS (continued)

Section Page

5.3.1 Robillard and Trahan's Lower Bound on $\bar{F}(t)$ 67

5.3.2 Kleindorfer's Upper and Lower Bounds on $\bar{F}(t)$ 69

5.3.3 Incorporating Different Methods of Approximating $\bar{F}(t)$ into Subnetwork Analysis 72

5.4 Additional Probability Inequalities as Bases for Upper and Lower Bounds on $\bar{F}(t)$ 74

6. CONCLUDING REMARKS 76

REFERENCES 77

LIST OF TABLES

Table		Page
1	Activity Durations for the Subnetwork in Figure 2	12
2	The Activity Durations for the Subnetwork in Figure 8	39
3	The Approximate Ordering of the x-values for the Subnetwork in Figure 8	40
4	Simulation Results for the Highly Skewed Left $F_{23,2}(t)$	46
5	Simulation Results for the Skewed Left $F_{8,2}(t)$	47
6	Simulation Results for the Symmetric $F_{5,5}(t)$	48
7	Simulation Results for the Skewed Right $F_{2,8}(t)$	49
8	Simulation Results for the Highly Skewed Right $F_{2,23}(t)$	50
9	Ratios of the Empirical Behavior of $G_2(t)$ Using Systematic Sampling to that Using Random Sampling	52
10	Extrapolation Data	59

LIST OF FIGURES

Figure		Page
1	A small project represented as a directed network	3
2	Subnetwork with activities labeled (activity number; mean duration)	12
3	Subnetwork for determining the associates of activity 2 when $\lambda = 1$	14
4	Subnetwork for determining the associates of activity 7 when $\lambda = 1$	14
5	Subnetwork for determining the associates of activity 9 when $\lambda = 1$	15
6	Subnetwork for determining the eliminants of activity 3 when $\theta = 2$	15
7	$G_1(t)$, $G_2(t)$, $G_3(t)$, $G_4(t)$, and $G_5(t)$ for the sample data in (3.7)	32
8	A small subnetwork	39
9	Duration distributions used in the simulation study	43
10	Extrapolation results for the data in Table 10	60
11	Subnetworks considered by Hartley and Wortham (1966)	64
12	Subnetworks considered by Ringer (1969)	65

1. A STATISTICAL APPROACH TO PROJECT SCHEDULING

1.1 Introduction

The many technological advances of the last century have resulted in a drastic increase in the magnitude and complexity of man's enterprises. This, in turn, has brought about an acute need for detailed and effective project planning. Thus, in recent years, a search for a general technique which can be employed to simplify the task of cost-effective project scheduling has been undertaken. A host of promising strategies have been proposed, and a few have even enjoyed widespread use. However, the methods currently in use have possibly serious shortcomings (see, for example, Sielken and Hartley (1977)). Therefore, under the sponsorship of the Office of Naval Research, the Institute of Statistics has undertaken the development, implementation, and evaluation of a new project scheduling system that yields reliable results and can be economically applied to very large scheduling problems. This report is a part of that undertaking.

1.2 The Project Scheduling Problem

Project scheduling problems arise in a wide variety of contexts. Consequently, a number of varying formulations of the problem are currently in use. Since these formulations are not all exactly equivalent, this subsection gives the specific formulation considered in this work.

A project is, in general, made up of a series of "tasks" or "activities" which consume time. These activities are represented graphically by directed arcs. The origin point and terminal point of an arc are both called "nodes". The graphical representation of a project, showing the precedence relationships among the various activities, is called a "network"; the first node in a network is usually referred to as the "source" while the last node is usually called the "sink". In addition, the following basic rules are adhered to:

1) Before a particular activity may begin, every other activity whose terminal node is that activity's origin node must be completed.

2) Arcs imply logical precedence only; the length of the arc has no significance.

3) The network cannot contain any loops or cycles.

For example, a small project might consist of activities A, B, C, D, and E with the following precedence relationships:

i) A must be completed before either C or D can be started;

ii) B must be completed before D can be started; and

iii) C and D must both be completed before E can be started.

The corresponding network representation is shown in Figure 1. The arc labeled F does not correspond to any "real" activity but is a "dummy" activity merely representing the precedence relation that A must be completed before D can be started. The circles numbered 1, 2, ..., 5 represent the activities' origin and terminal nodes.

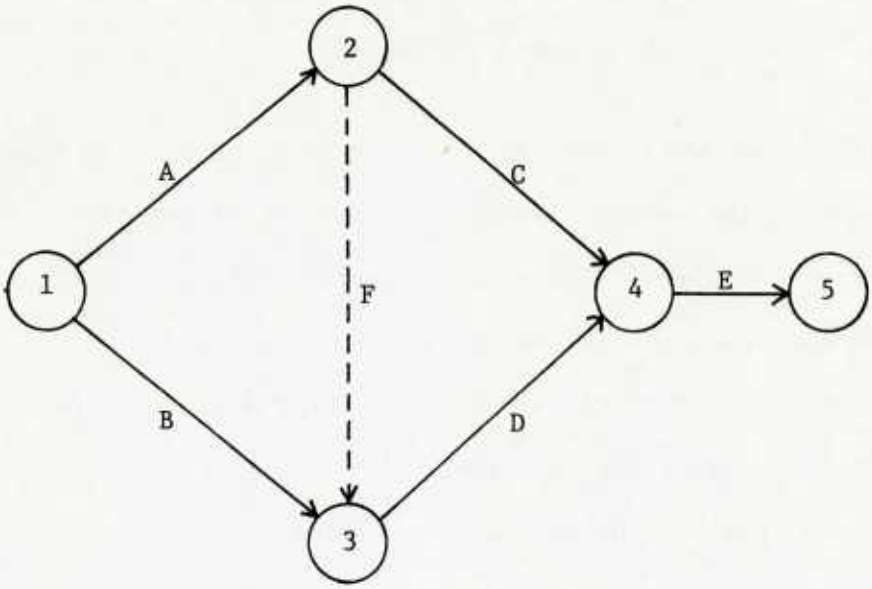


Figure 1

A small project represented as a directed network.

The time required to complete an activity is a random variable. The cost of an activity is a convex piecewise linear function of the activity's mean duration. Thus, a "project schedule" is a specification of each activity's mean duration. The "total project cost" is simply the sum of the corresponding activity costs. The time to complete the entire project is a random variable whose distribution depends upon the activity duration distributions. The objective is to determine a minimum cost project schedule such that the mean (or some percentile) of the corresponding project completion time distribution is less than or equal to a specified project deadline.

1.3 Outline of the New Approach to Project Scheduling

In 1974 the development of a new approach to project scheduling was begun with the support of the Office of Naval Research. The new project scheduling procedure that has resulted is an iterative algorithm involving the following five general steps:

Step 1. Deterministic Scheduling: Find a minimum cost project schedule which completes the project by TARGET TIME when each activity's duration is exactly its mean duration and hence deterministic instead of random. (The initial value of TARGET TIME is usually the specified project deadline.)

Step 2. Simplification: Let each activity's duration be a random variable with distribution corresponding to that activity's mean duration chosen during Deterministic Scheduling. Replace various configurations of activities by single activities. The duration distribution for a replacement activity is the distribution of the time to complete all of the activities in the configuration it is replacing. The result of this step is a simplified project network with fewer activities.

Step 3. Decomposition: Partition the simplified project network into several subnetworks in such a way that the resultant subnetworks can be linked together in

either series or parallel to form the simplified project network.

Step 4. Subnetwork Analysis: Analyze separately each of the subnetworks determined during Decomposition. Within a subnetwork each activity's duration distribution is approximated by a two-point discrete distribution with matching mean, variance, and third moment. Determine the subnetwork duration distribution corresponding to these discrete activity duration distributions.

Step 5. Synthesis: Combine the approximate subnetwork duration distributions to obtain an approximate completion time distribution for the entire project. If the mean (or some specified percentile), $\hat{T}$, of this project completion time distribution is sufficiently close to the specified project deadline, the "optimal" project schedule has been found. Otherwise, reset TARGET TIME to

New TARGET TIME = Old TARGET TIME*(Project Deadline/$\hat{T}$)

and return to Step 1.

A general discussion and relatively nonmathematical overview of this project scheduling procedure is contained in Baker and Sielken (1978)(see also Sielken and Hartley (1977)). The detailed documentation of the development thus far of each step is as follows:

Step 1. Dunn and Sielken (1977);

Step 2. Hartley and Wortham (1966) and Ringer (1969);

Step 3. Sielken and Fisher (1976);

Step 4. Sielken, Ringer, Hartley, and Arseven (1974) and Sielken, Hartley, and Spoeri (1976);

Step 5. Sielken, Ringer, Hartley, and Arseven (1974) and Sielken, Hartley, and Spoeri (1976).

1.4 An Integrated System of Computer Programs

Prior to this research, separate computer programs had been written to perform each of the five steps. (These programs are fully documented in the references cited for each step.) However, from a user's viewpoint, this arrangement was awkward because the programs had to be executed one at a time and the output from each step had to be manually modified for use by the next program in the sequence. Thus, one of the objectives of this research has been to fashion the individual programs into an integrated package that is more practicable from a user's viewpoint.

The ability to schedule large projects with as many as 1000 activities in the project network and as many as 500 activities in any simplified subnetwork was one of the desired characteristics of the computer implementation. This objective prohibited the combination of the five original programs into a single large program since doing so would drastically limit the size of the project that could be analyzed because of the computer core storage restrictions. Thus, the computer implementation of the new project scheduling procedure has been in the form of several individual programs internally linked

together. The resulting integrated system of computer programs requires that the user supply the project description and algorithm parameters only to the main (first) program. From then on each program automatically prepares the proper input for the remaining programs in the iterative procedure and stores this information on either disk or tape from which it is retrieved as needed. The job control language automatically calls the individual programs as it cycles through the five step iterative algorithm.

Since the new project scheduling procedure is an iterative procedure, it may repeat Steps 1-5. However, as pointed out in Sielken and Hartley (1977), after the initial performance of Steps 1-3 their subsequent performance is greatly simplified. Thus, special simplified versions of the programs for these steps are called when these steps are repeated. Needless to say, the preparation of these simplified versions has greatly improved the efficiency of the computer implementation.

The new project scheduling software package that has resulted is fully documented in the User's Guide found in Baker and Sielken (1978). Included in Baker and Sielken (1978) is an example with a complete listing of the system's input and output.

1.5 The Determination of the Subnetwork Duration Distribution

Two new theoretical contributions to the project scheduling procedure are documented in this report. Both are improvements to the statistical methodologies used in determining the subnetwork

duration distributions. Section 2 of this report contains a detailed description of the procedure (including the improvements) used to determine the subnetwork duration distributions. Sections 3 and 4 present detailed documentations of the improvements.

From a statistical viewpoint a subnetwork's duration is defined easily enough as the maximum of the paths through the subnetwork. In Figure 1 (p. 3), for example, there are really three paths; namely,

$$P_1 = A + C + E$$

$$P_2 = A + F + D + E$$

$$P_3 = B + D + E \ . \qquad (1.1)$$

The project duration is simply the maximum of P_1, P_2, and P_3. However, the difficulty is that the paths are usually dependent since the paths often have activities in common. For example, the paths P_1 and P_2 have activities A and E in common. Section 5 contains a review of the few known general results concerning the distribution of the maximum of dependent random variables. Also in Section 5 is an indication of how these general results could be used to modify the Subnetwork Analysis procedure described in Section 2.

2. ANALYSIS OF A SUBNETWORK

2.1 Introduction

The objective of Subnetwork Analysis is to determine each subnetwork's duration distribution.

At the end of Step 2 each activity in the subnetwork has a specified duration distribution. This distribution is now approximated by a two-point discrete distribution. In particular, an activity, say A, is now conceptualized as having two possible duration times, say ℓ_A for a lower duration and u_A for an upper duration. The probability that the activity duration is ℓ_A is assumed to be P_A, and correspondingly the probability that the activity duration is u_A is assumed to be $Q_A = 1 - P_A$. The values of ℓ_A, u_A, and P_A are chosen so that the mean, variance, and third moment of the discrete distribution are the same as the mean, variance, and third moment of activity A's specified duration distribution.

Let n be the number of activities in the subnetwork. Let $v = 1, 2, \ldots, 2^n$ index the 2^n possible configurations of activity durations when each activity is either at its upper duration or at its lower duration. Let

$$p_v = \text{probability of the v-th activity duration configuration}$$

$$= \prod_{i=1}^{n} [P_i(1 - \delta_{v,i}) + Q_i\delta_{v,i}] \qquad (2.1)$$

where

$\delta_{v,i} = 1$ if the duration for the i-th activity is u_i in the v-th activity duration configuration

$= 0$ if the duration for the i-th activity is ℓ_i in the v-th activity duration configuration. (2.2)

Then the subnetwork duration distribution when each activity has its two-point discrete distribution is

$$F(t) = \sum_{v=1}^{2^n} p_v I_t(t_v) \tag{2.3}$$

where

t_v = the subnetwork duration when the activity durations are in the v-th configuration (2.4)

and

$$I_t(t_v) = 1 \quad \text{if } t_v \leq t ,$$
$$= 0 \quad \text{if } t_v > t . \tag{2.5}$$

The discrete distribution function F is an approximation to the subnetwork's exact duration distribution.

The goal of Subnetwork Analysis is to determine F.

Since the number, n, of activities in the subnetwork may be fairly large, the complete enumeration of the 2^n discrete subnetwork durations may sometimes be impractical. When this happens, the discrete subnetwork duration distribution F must be approximated. The approximation of F will be based on the activities which are mostly likely to influence the subnetwork duration. The identification of these important activities and their interrelationships is discussed in the next subsection which is a review of the procedures originating in Sielken,

Ringer, Hartley, and Arseven (1974) and Sielken, Hartley, and Spoeri (1976).

Each subnetwork is assumed to be an acyclic network with one source, one sink, and no cut vertices.

2.2 Formation of Clusters

The mean duration for activity A is

$$m_A = P_A \ell_A + Q_A u_A \ , \tag{2.6}$$

and the standard deviation of activity A's duration is

$$s_A = [P_A \ell_A^2 + Q_A u_A^2 - m_A^2]^{\frac{1}{2}} \ . \tag{2.7}$$

When each activity duration takes on a fixed (nonrandom) value, the subnetwork's duration is the duration of the longest path through the subnetwork where the "length" of an activity is its duration. For example, consider the subnetwork described in Table 1 and displayed in Figure 2. When each activity duration is its mean duration, then the subnetwork's duration is 32, corresponding to the path consisting of activities 2, 7, and 9.

Definition 1: A critical activity is an activity on the longest path when all the subnetwork's activity durations are set to their means.

Thus in the example the critical activities are 2, 7, and 9.

The search for the activities which are most likely to influence the subnetwork duration begins with the critical activities. Each critical activity initiates a separate set of activities called a

TABLE 1

Activity Durations for the Subnetwork in Figure 2

Activity	ℓ_A	u_A	P_A	m_A	s_A
1	0.00	2.00	.5	1	1
2	8.00	10.50	.2	10	1
3	9.55	15.67	.6	12	3
4	1.50	4.00	.8	2	1
5	3.35	5.52	.7	4	1
6	4.00	6.00	.5	5	1
7	8.73	16.90	.6	12	4
8	12.00	14.50	.2	14	1
9	5.00	15.00	.5	10	5

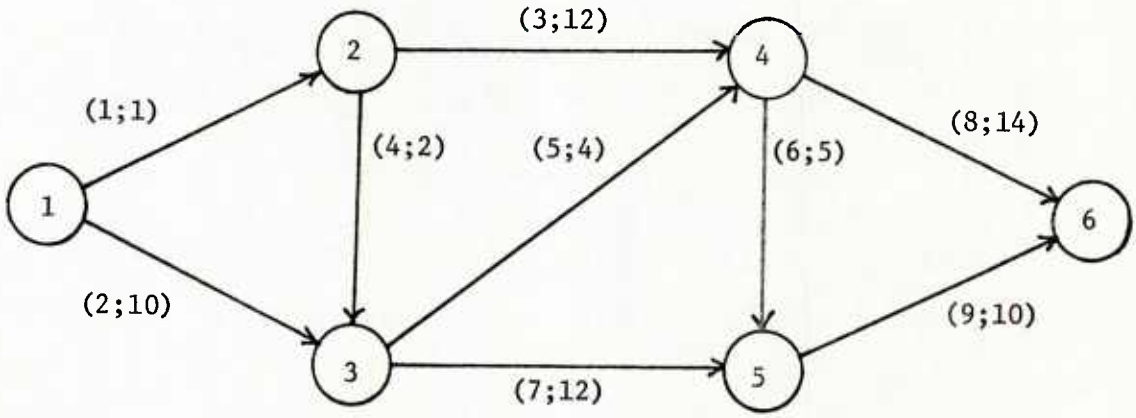


Figure 2

Subnetwork with activities labeled (activity number; mean duration).

"cluster". Initially there are several clusters. In the example the initial clusters are

$$C_1 = \{2\}, \quad C_2 = \{7\}, \quad \text{and } C_3 = \{9\}. \tag{2.8}$$

Some of the non-critical activities may influence the subnetwork's duration when not all of the activity durations are at their mean values.

Definition 2: An associate of a critical activity A is a non-critical activity which is on the longest path when all activity durations are set to their mean except for the critical activity A which has its duration reduced from m_A to $\max(m_A - \lambda s_A, 0)$ where λ is a nonnegative parameter.

Thus the associates of a critical activity A are those activities whose effect on the subnetwork's duration are related to activity A's duration. In the example, for $\lambda = 1$ the associates of the critical activities 2, 7, and 9 can be determined by considering Figures 3, 4, and 5 respectively. In Figure 3 the longest path is still the critical path 2, 7, and 9, so that activity 2 has no associates. In Figure 4 the longest path is 2, 5, 6, and 9, so that activities 5 and 6 are the associates of activity 7. In Figure 5, the longest path is 2, 5, and 8, so that activities 5 and 8 are associates of activity 9.

The associates of each critical activity are determined and added to the cluster containing that critical activity. Thus, in the example the clusters are expanded to

$$C_1 = \{2\}, \quad C_2 = \{7, 5, 6\}, \quad \text{and } C_3 = \{9, 5, 8\}. \tag{2.9}$$

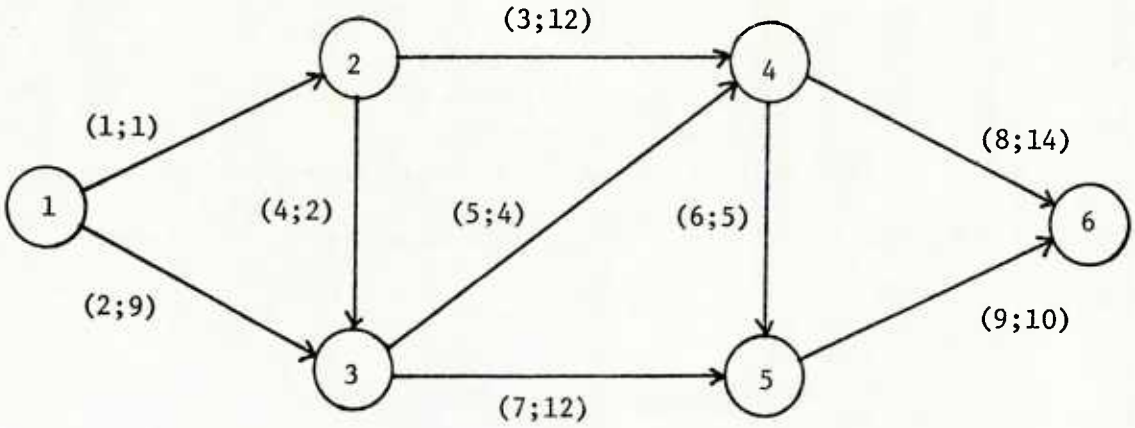


Figure 3

Subnetwork for determining the associates of Activity 2 when $\lambda = 1$.

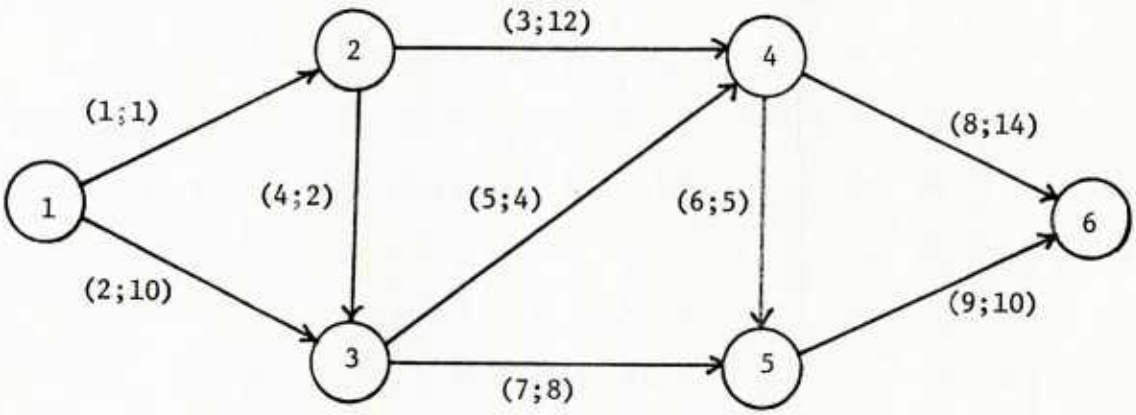


Figure 4

Subnetwork for determining the associates of Activity 7 when $\lambda = 1$.

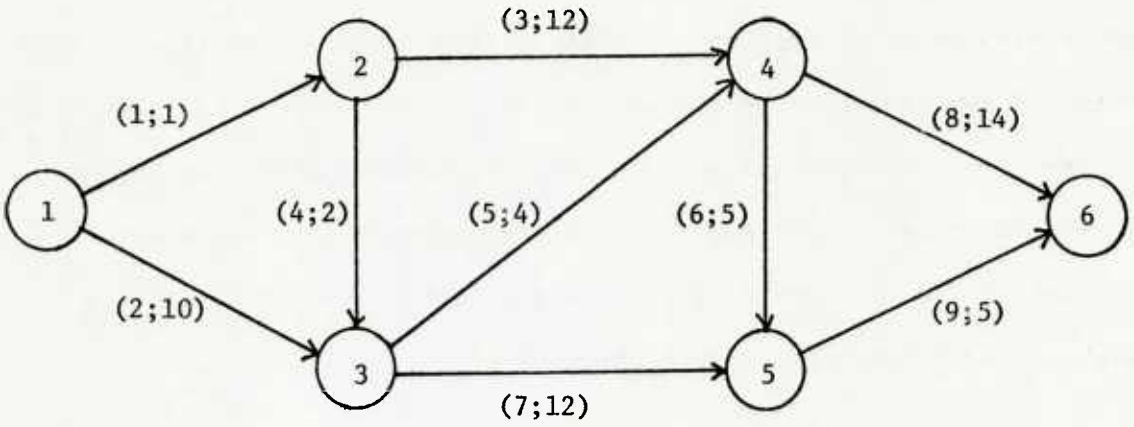


Figure 5

Subnetwork for determining the associates of Activity 9 when $\lambda = 1$.

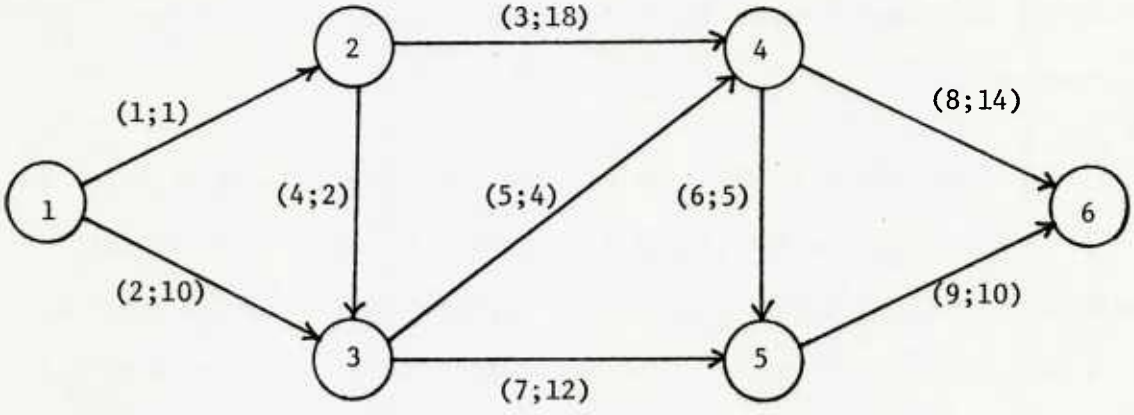


Figure 6

Subnetwork for determining the eliminants of Activity 3 when $\theta = 2$.

The idea underlying the clusters is that they should be sets of activities whose effects on the subnetwork's duration are interrelated. Thus, if two clusters contain any activities in common, the activities in these two clusters all have an interrelated effect on the subnetwork's duration, so the two clusters are combined into one cluster. In the example clusters C_2 and C_3 both contain activity 5, so they are combined. The resulting clusters are

$$C_1 = \{2\} \text{ and } C_2 = \{5, 6, 7, 8, 9\} . \tag{2.10}$$

A non-critical activity may also influence the subnetwork's duration if its duration exceeds its mean.

Definition 3: An eliminant of a non-critical activity A is a critical activity which is not on the longest path when all activity durations are set to their means except for activity A which has its duration increased from m_A to $m_A + \theta s_A$ where θ is a nonnegative parameter.

For instance, if $\theta = 2$, the eliminants of the non-critical activity 3 in the example can be determined from Figure 6. There the longest path is 1, 3, 6, and 9, so that the eliminants of activity 3 are the critical activities 2 and 7. In the example, when $\theta = 2$, none of the other non-critical activities (1, 4, 5, 6, and 8) have any eliminants. For a specified value of θ the eliminants of every non-critical activity are determined. If a non-critical activity A has eliminants, then the effect of A's eliminants on the subnetwork duration is related to A's duration, so A is added to every cluster containing at least one

of its eliminants. Thus in the example the clusters become

$$C_1 = \{2, 3\} \text{ and } C_2 = \{3, 5, 6, 7, 8, 9\} . \qquad (2.11)$$

After the clusters have been expanded on the basis of eliminants, any two clusters containing common elements are combined. Therefore in the example, C_1 and C_2 are combined to form a single cluster

$$C_1 = \{2, 3, 5, 6, 7, 8, 9\} . \qquad (2.12)$$

In general, after the determination of associates and eliminants for specified values of λ and θ and the subsequent combining of clusters, there may still be more than one cluster and some of the non-critical activities may not be in any cluster. Usually the larger the values of λ and θ the greater the number of activities in the clusters and the smaller the number of clusters. The clusters that remain represent sets of activities such that the effects on the subnetwork's duration of the activity durations for the activities within a set are all interrelated. Activities in different clusters have roughly independent effects on the subnetwork's duration. Activities not in any cluster have essentially no effect on the subnetwork's duration.

The consideration of critical activities, associates, eliminants, and the formation of clusters of related activities is obviously only one way of identifying the activities which have an important effect on the subnetwork's duration and their interrelationships. However, this particular procedure does have the following desirable properties:

Property 1: If $\lambda_2 > \lambda_1$, then any activity which would be an associate of a critical activity A when $\lambda = \lambda_1$ would also be an associate of A when $\lambda = \lambda_2$.

Property 2: If $\theta_2 > \theta_1$, then any critical activity which would be an eliminant of a non-critical activity A when $\theta = \theta_1$ would also be an eliminant of A when $\theta = \theta_2$.

Property 3: For any fixed value of λ, the set of activities in the union of the clusters is monotically nondecreasing as $\theta \to \infty$.

Property 4: The number of clusters is nonincreasing as $\theta \to \infty$.

Property 5: If $s_A > 0$ for a non-critical activity A, then there exists $\theta_A < \infty$ such that A will have some eliminants for any $\theta \geq \theta_A$.

Property 6: If $s_A > 0$ for every non-critical activity A and

$$\theta^* = \max\{\theta_A;\ A \text{ non-critical}\} ,$$

then for $\theta \geq \theta^*$ all activities will be in one cluster.

Most of these properties are fairly straightforward; however, Property 6 requires some special justification. This justification is based on the following definition and theorem which is proven in Sielken, Ringer, Hartley, and Arseven (1974).

Definition 4: In any acyclic network a bridge over any two consecutive arcs A_1 and A_2 is any arc A_3 such that all paths from the source to the sink passing through A_3 do not pass through either A_1 or A_2.

Theorem 1: In any acyclic network with no cut vertices there is at least one bridge for any pair of consecutive arcs.

Property 5 implies that all activities will belong to some cluster if $\theta \geq \theta^*$. Now consider any two consecutive activities A_1 and A_2 on the critical path. Theorem 1 implies that there is a bridge over A_1 and A_2, say A_3. Since the critical path passes through A_1 and A_2, A_3 cannot be on the critical path. Therefore, if $\theta \geq \theta^* \geq \theta_{A_3}$, A_1 and A_2 will be eliminants of A_3 and hence will be in the same cluster as A_3. Thus, since each cluster contains at least one original critical activity and any two consecutive critical path activities belong to the same cluster when $\theta \geq \theta^*$, there is only one cluster when $\theta \geq \theta^*$ and Property 6 is established.

2.3 Bounding the Discrete Subnetwork Duration Distribution F

2.3.1 Upper Bounds on F

Suppose that the cluster formation procedure described in subsection 2.2 has been carried out on a subnetwork for some specified values of θ and λ and yielded K clusters. For each cluster C so determined, let n_c be the number of activities in the cluster and let $v = 1, \ldots, 2^{n_c}$ index the 2^{n_c} configurations of activity durations corresponding to

(a) the duration for each activity A not in C being equal to its lower point ℓ_A, and

(b) the durations for the activities in C being at each of the 2^{n_c} possible combinations of their upper and lower points.

Then define

$$F^{+}(C;\ t) = \sum_{v=1}^{2^{n_C}} p_v I_t(t_v) \tag{2.13}$$

where p_v, t_v, and $I_t(t_v)$ are defined in (2.1), (2.4), and (2.5) respectively. The distribution function $F^{+}(C;\ t)$ is an upper bound on F. This can be shown by considering the following:

Theorem 2: For any cluster C, any t, and any activity A not in C,

$$F^{+}(C \cup \{A\};\ t) \leq F^{+}(C;\ t) .$$

(For the proof of this theorem, see Sielken, Hartley, and Spoeri (1975).) A straightforward application of Theorem 2 yields

Theorem 3: For any two clusters C_1 and C_2 and any t,

$$F^{+}(C_1 \cup C_2;\ t) \leq \min\{F^{+}(C_1;\ t),\ F^{+}(C_2;\ t)\} .$$

If C* represents the set (cluster) of all activities in the subnetwork, then

$$F(t) = F^{+}(C^*;\ t) . \tag{2.14}$$

Since C is a subset of C*, either Theorem 2 or Theorem 3 implies

$$F^{+}(C;\ t) \geq F(t) \tag{2.15}$$

for any cluster C.

Theorems 2 and 3 can also be used to define some tighter upper bounds on the subnetwork's duration distribution than $F^{+}(C;\ t)$. Historically, two different improved bounds have been employed, and both have been incorporated into the current subnetwork analysis

procedure. They are

$$F_1^+(t;\ \theta,\ \lambda) = F^+\Big(\bigcup_{i=1}^{K} C_i;\ t\Big) \tag{2.16}$$

and

$$F_2^+(t;\ \theta,\ \lambda) = \min_{1\le i\le K} F^+(C_i;\ t)\ . \tag{2.17}$$

Let $F^+(t;\ \theta,\ \lambda)$ denote either $F_1^+(t;\ \theta,\ \lambda)$ or $F_2^+(t;\ \theta,\ \lambda)$. Then, since Property 2 of the cluster formation procedure implies that as θ increases the clusters expand or are combined, Theorems 2 and 3 imply that $F^+(t;\ \theta,\ \lambda)$ is a nonincreasing function of θ for every t and any λ. Property 6 and (2.14) imply that for $\theta \ge \theta^*$

$$F^+(t;\ \theta,\ \lambda) = F(t) \tag{2.18}$$

for every t and any λ. Also (2.14) along with the definitions (2.16) and (2.17) imply

$$F^+(t;\ \theta,\ \lambda) \ge F(t) \tag{2.19}$$

for all t, θ, and λ. These results are summarized in Theorem 4.

Theorem 4: (a) $F^+(t;\ \theta,\ \lambda)$ is a nonincreasing function of θ for every t and any λ;

(b) there exists a finite value θ^* such that $\theta \ge \theta^*$ implies $F^+(t;\ \theta,\ \lambda) = F(t)$ for every t and λ; and

(c) for any θ, λ, and t

$$F^+(t;\ \theta,\ \lambda) \ge F(t)\ .$$

2.3.2 Lower Bounds on F

Let n_c denote the number of activities in cluster C, and let $v = 1, \ldots, 2^{n_c}$ index the 2^{n_c} configuration of activity durations corresponding to

(a) the duration for each activity A not in the cluster being equal to its upper point u_A, and

(b) the durations for activities in the cluster being at each of the 2^{n_c} possible combinations of the upper and lower points.

Then define

$$F^{-}(C;\ t) = \sum_{v=1}^{2^{n_c}} p_v I_t(t_v) \tag{2.20}$$

where p_v, t_v, and $I_t(t_v)$ are as previously defined. Also define

$$F_1^{-}(t;\ \theta,\ \lambda) = F^{-}\left(\bigcup_{i=1}^{K} C_i;\ t\right) \tag{2.21}$$

and

$$F_2^{-}(t;\ \theta,\ \lambda) = \max_{1 \le i \le K} F^{-}(C_i;\ t)\ . \tag{2.22}$$

Using an argument completely analagous to that used to prove Theorem 4, Sielken, Hartley, and Spoeri (1975) also proved

Theorem 5: (a) $F^{-}(t;\ \theta,\ \lambda)$ is a nondecreasing function of θ for any fixed value of λ;

(b) there exists a finite value θ^* such that $\theta \ge \theta^*$ implies

$$F^{-}(t;\ \theta,\ \lambda) = F(t)$$

for every t and any λ; and

(c) for any θ, λ, and t

$$\bar{F}(t;\ \theta,\ \lambda) \leq F(t)\ .$$

(Again, $\bar{F}(t;\ \theta,\ \lambda)$ is a generic term used to denote either $\bar{F}_1(t;\ \theta,\ \lambda)$ or $\bar{F}_2(t;\ \theta,\ \lambda)$.) Thus, $\bar{F}(t;\ \theta,\ \lambda)$ is a valid lower bound on F.

2.3.3 The Tightness of the Bounds on F

That the F_1-bounds are tighter than the F_2-bounds can be seen as follows. The evaluation of $\bar{F}_2(t;\ \theta,\ \lambda)$ involves the determination of $\bar{F}(C_i;\ t)$ for each i whereas $\bar{F}_1(t;\ \theta,\ \lambda) = \bar{F}(\overset{K}{\underset{i=1}{U}} C_i,\ t)$. Let L_i be the length of the longest path when

1) the activities in C_i are at a particular configuration of their upper and lower durations and
2) all activities not in C_i have their upper durations.

Let L_U be the length of the longest path when

1) the configuration of upper and lower durations for the activities in C_i is the same as in the determination of L_i,
2) the activities in $\overset{K}{\underset{j=1}{U}} C_j - C_i$ are at <u>any</u> combination of their upper and lower durations, and
3) all activities not in $\overset{K}{\underset{j=1}{U}} C_j$ have their upper durations.

Then $L_i \geq L_U$ since every activity duration in the determination of L_i is greater than or equal to its corresponding duration in the determination of L_U. Since $L_i \geq L_U$ for any configuration of upper and lower durations for the activities in C_i,

$$\bar{F}(\overset{K}{\underset{j=1}{U}} C_j;\ t) \geq \bar{F}(C_i;\ t) \qquad (2.23)$$

and

$$F_1^-(t;\theta,\lambda) = F^-\left(\bigcup_{j=1}^{K} C_j;\ t\right) \geq \max_{1\leq i\leq K} F^-(C_i;\ t) = F_2^-(t;\theta,\lambda)\ . \qquad (2.24)$$

A similar argument can be used to show

$$F_1^+(t;\theta,\lambda) = F^+\left(\bigcup_{j=1}^{K} C_j;\ t\right) \leq \min_{1\leq i\leq K} F^+(C_i;\ t) = F_2^+(t;\theta,\lambda)\ . \qquad (2.25)$$

The extent of the differences between the two upper bounds and two lower bounds depends heavily on the structure of the particular subnetwork being analyzed and is a topic that should be considered in future empirical studies.

2.4 Using Sampling to Estimate the Upper and Lower Bounds on F

The only instance in which upper and lower bounds on F are computed rather than F itself is when it is computationally impractical to determine the longest path for each of the 2^n activity duration configurations.

For given θ and λ, the evaluation of $F_1^+(t;\theta,\lambda)$ only requires the determination of the longest path for each of 2^{n_U} activity configurations where n_U is the number of activities in the union of the clusters $C_o = \bigcup_{j=1}^{K} C_j$; i.e.,

$$n_U = \sum_{j=1}^{K} n_j\ . \qquad (2.26)$$

The evaluation of $F_1^-(t;\theta,\lambda)$ also entails only 2^{n_U} longest path determinations. Likewise, the evaluation of $F_2^+(t;\theta,\lambda)$ or $F_2^-(t;\theta,\lambda)$ only requires the determination of the longest path for each of

$$n_s = \sum_{i=1}^{K} 2^{n_i} \qquad (2.27)$$

activity configurations. Since 2^{n_U} is always greater than or equal to

n_s, $F_2^+(t;\ \theta,\ \lambda)$ and $F_2^-(t;\ \theta,\ \lambda)$ are the most economical bounds to compute in terms of the number of longest path determinations required. However, for any given θ and λ, $F_1^+(t;\ \theta,\ \lambda)$ and $F_1^-(t;\ \theta,\ \lambda)$ are tighter bounds than $F_2^+(t;\ \theta,\ \lambda)$ and $F_2^-(t;\ \theta,\ \lambda)$, respectively. Thus, in making the choice of which one of the two sets of bounds to compute, there is a trade-off between the accuracy of the bounds and the effort required to compute them.

Since the cluster formation procedure is such that the clusters expand or are pooled as θ increases, it may happen that for particular θ and λ, 2^{n_U} and 2^{n_i} for some i are both quite large even though θ is only moderately large. In this case it again becomes impractical to examine <u>all</u> the required activity configurations involved in determining either the F_1-bounds or the F_2-bounds. Consequently, if for the specified values of θ and λ, 2^{n_U} (or 2^{n_i} for some i, as the case may be) is excessively large, Subnetwork Analysis will compute estimates of the corresponding upper and lower bounds based on only a sample of the total number of possible configurations. The actual estimators used in this situation are described and developed in Section 3.

2.5 Estimating F by Extrapolating Between the Upper and Lower Bounds on F

Theorems 4 and 5 of subsection 2.3 imply that if $\theta_{i+1} \geq \theta_i$ and $\lambda_{i+1} \geq \lambda_i$ for all i = 1, ..., I then

$$F^+(t;\ \theta_1,\ \lambda_1) \geq F^+(t;\ \theta_2,\ \lambda_2) \geq \ldots \geq F^+(t;\ \theta_I,\ \lambda_I) \geq F(t) \geq$$

$$F^-(t;\ \theta_I,\ \lambda_I) \geq F^-(t;\ \theta_{I-1},\ \lambda_{I-1}) \geq \ldots \geq F^-(t;\ \theta_1,\ \lambda_1) \qquad (2.28)$$

for all t. Thus, if $F^{+}(t; \theta, \lambda)$ and $F^{-}(t; \theta, \lambda)$ have been calculated for I pairs (θ_i, λ_i) i = 1, ..., I $(\theta_{i+1} \geq \theta_i, \lambda_{i+1} \geq \lambda_i)$, then F(t) may be estimated by extrapolating between $F^{+}(t; \theta_I, \lambda_I)$ and $F^{-}(t; \theta_I, \lambda_I)$. As currently written, Subnetwork Analysis calculates upper and lower bounds on the subnetwork's approximate duration distribution for a sequence of three (θ, λ) pairs, (θ, λ) = (1, 1), (2, 2), (3, 2). An extrapolation procedure is then used to obtain an estimate of F. The procedure that has been developed for this purpose is documented in Section 4 of this report.

2.6 A Summary of the Subnetwork Analysis Procedure

The following is a step-by-step description of the subnetwork analysis procedure in summary form. Recall that the objective of Subnetwork Analysis is to determine an "approximation", say $\hat{F}$, to the subnetwork's duration distribution.

(a) If n = 1, let $\hat{F}$ be the actual activity duration distribution for the one activity comprising the subnetwork, and stop. Otherwise, go to Step b.

(b) Identify the two-point discrete distribution (ℓ_A, u_A, P_A, Q_A) for every activity A in the subnetwork.

(c) Ascertain the user's choice of

(1) NMAX, the maximum value of m for which all 2^m activity duration configurations are to be explicitly considered,

(2) the (θ, λ) pairs to be considered if not the standard pairs (1, 1), (2, 2), and (3, 2),

(3) whether the bounds on F are to be (F_1^-, F_1^+) or (F_2^-, F_2^+) if n > NMAX, and

(4) SAMSIZ, the sample size to be taken if, in the determination of bounds on F for some (θ, λ) pair, the number of activity configurations in the cluster being considered exceeds 2^{NMAX}.

(d) If the number of activities in the subnetwork doesn't exceed NMAX, compute the subnetwork's discrete duration distribution, F, explicitly, let $\hat{F} = F$, and stop. Otherwise, go to Step e.

(e) Do Steps f - i for every (θ, λ) pair. Then go to Step j.

(f) Form the clusters corresponding to (θ, λ). If the bounds are to be (F_1^-, F_1^+), go to Step g. If the bounds are to be (F_2^-, F_2^+), go to Step h.

(g) Form the union of the clusters and determine n_U. If $n_U \leq$ NMAX, evaluate the bounds (F_1^-, F_1^+) on the basis of all 2^{n_U} activity duration configurations. If $n_U >$ NMAX, take a sample of size SAMSIZ from the 2^{n_U} activity duration configurations and form both F_1^- and F_1^+ on the basis of this single sample. Go to Step e.

(h) Do the following for each cluster, C_i. Let n_i denote the number of activities in the cluster. If $n_i \leq$ NMAX, evaluate $F^-(C_i; t)$ and $F^+(C_i; t)$ on the basis of all 2^{n_i} activity duration configurations. If $n_i >$ NMAX, take a sample of size SAMSIZ from the 2^{n_i} activity duration configurations and form both $F^-(C_i; t)$ and $F^+(C_i; t)$ on the basis of this single sample.

(i) Form F_2^- and F_2^+ from the $F^-(C_i; t)$'s and $F^+(C_i; t)$'s respectively. Go to Step e.

(j) Form $\hat{F}$ by extrapolating the (F^-, F^+) bounds determined for the (θ, λ) pairs. Stop.

This process is repeated for every subnetwork in the simplified project network.

3. SAMPLE-BASED ESTIMATORS FOR A DISCRETE DISTRIBUTION FUNCTION

3.1 Introduction

In the subnetwork analysis procedure the calculation of $F^{-}(C; t)$ or $F^{+}(C; t)$, say $F(C; t)$, for a cluster C comprised of n_c activities requires 2^{n_c} longest path determinations. If 2^{n_c} is too large from a practical standpoint, $F(C; t)$ must be approximated on the basis of a sample of the possible 2^{n_c} activity duration configurations.

The estimation of $F(C; t)$ involves two aspects

1) the identification of an acceptable method of sample selection, and

2) the determination of the form of the estimator.

Because of the practical difficulties (computer storage requirements, etc.) involved in implementing other sampling schemes, only simple random sampling (with replacement) and systematic sampling were considered in this research. Of these two, systematic sampling is the preferred technique. The reasons for this preference are presented in subsection 3.4.

Now $F(C; t)$ is the distribution function of a discrete random variable X (the length of the subnetwork's longest path) which has a known number,

$$M = 2^{n_c} \tag{3.1}$$

of possible values which are not necessarily distinct. Since the probability that a particular activity, A, attains its lower duration is known to be P_A and the probability that it attains its upper

duration is known to be Q_A, the random variable X is such that when an activity duration configuration is realized not only is the numerical value of X observed but also the probability, p, of that activity duration configuration is available. This departure from the usual estimation situation has been exploited in the formation of the estimators for F(C; t).

3.2 Some Proposed Estimators

The five estimators of F(C; t) that were considered in this research are

$$G_1(t) = \frac{\sum_{i=1}^{m} I_t(x_i)}{m} ; \tag{3.2}$$

$$G_2(t) = \frac{\sum_{i=1}^{m} p_i I_t(x_i)}{\sum_{i=1}^{m} p_i} ; \tag{3.3}$$

$$G_3(t) = \begin{cases} 0, & t < x_1 \\ \dfrac{\sum_{i=1}^{m} p_i I_t(x_i) + \dfrac{(t-x_j)p_{j+1}}{(x_{j+1}-x_j)}}{\sum_{i=1}^{m} p_i}, & x_1 \leq t < x_m \\ 1, & t \geq x_m ; \end{cases} \tag{3.4}$$

$$G_4(t) = \begin{cases} 0, & t < x_1 \\ \dfrac{\sum_{i=1}^{m} p_i I_t(x_i) + \dfrac{(t-x_j)p_{j+1}(1-\sum_{i=1}^{m} p_i)}{(x_{j+1}-x_j)}}{\sum_{i=1}^{m} p_i}, & x_1 \leq t < x_m ; \\ 1, & t \geq x_m \end{cases} \tag{3.5}$$

and

$$G_5(t) = \begin{cases} 0, & t < x_1 \\ \sum_{i=1}^{m} p_i I_t(x_i) + \dfrac{(t-x_1)(1-\sum_{i=1}^{m} p_i)}{(x_m-x_1)}, & x_1 \leq t < x_m ; \\ 1, & t \geq x_m \end{cases} \tag{3.6}$$

where in each case, the x_i's represent an ordered sample of size m from the population of subnetwork durations corresponding to the M activity duration configurations, p_i is the probability of the activity duration configuration corresponding to x_i, j is the largest integer such that $x_j \leq t$ and $I_t(\cdot)$ is as defined in (2.5).

Even though sampling will normally only be employed if M is very large, for illustration purposes consider M = 20 and that the following ordered sample of size m = 5 has been obtained:

$$\begin{aligned} &x_1 = 2,\ x_2 = 2.5,\ x_3 = 4.0,\ x_4 = 4.5,\ x_5 = 5 \\ &p_1 = .02,\ p_2 = .03,\ p_3 = .05,\ p_4 = .07,\ p_5 = .03\ . \end{aligned} \tag{3.7}$$

The $G_1(t)$, $G_2(t)$, ..., $G_5(t)$ for this sample data are displayed in Figures 7a - 7e, respectively.

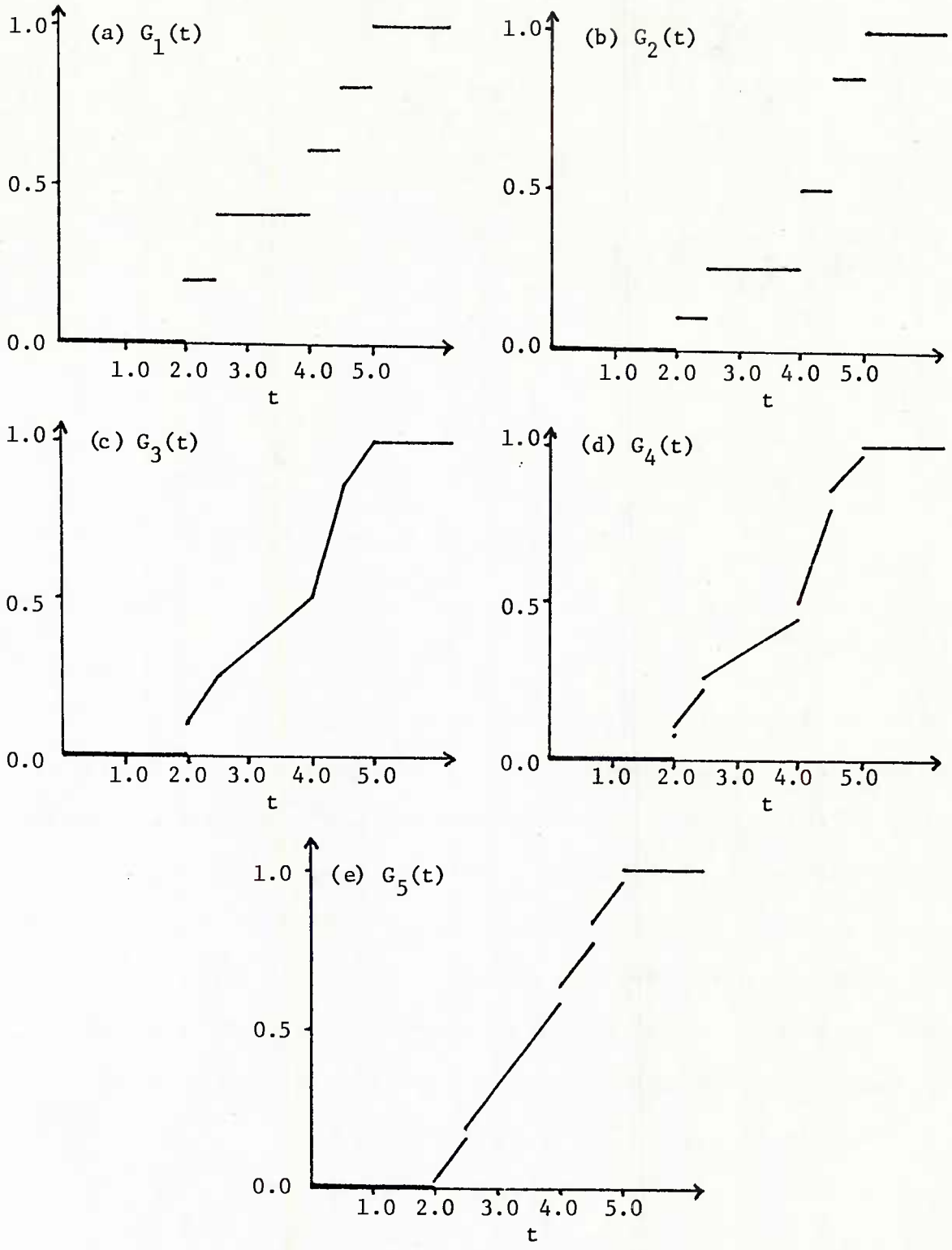


Figure 7

$G_1(t)$, $G_2(t)$, $G_3(t)$, $G_4(t)$, and $G_5(t)$ for the sample data in (3.7).

The discussion in subsections 3.2.1 - 3.2.5 assumes that the m sample points have been selected without replacement.

3.2.1 The Empirical Distribution Function, $G_1(t)$

In many situations requiring the estimation of a distribution function, the empirical distribution function, $G_1(t)$, has very desirable properties. However, these properties are derived from the assumption that the sample values have been observed with roughly the same relative frequency as they occur in the population. For the situation under consideration, though, every x_i occurs in the sample with relative frequency $\frac{m}{M}$ regardless of the true value of p_i. Consequently, $G_1(t)$ was of interest in this study only as a basis for comparison and generalization.

3.2.2 The Modified Empirical Distribution Function, $G_2(t)$

The major disadvantage of $G_1(t)$ is that it ignores the information contained in the p_i's. Instead of assigning weight $\frac{1}{m}$ to each sampled point, $G_2(t)$ assigns to x_i the weight

$$p_i / \sum_{i=1}^{m} p_i \ . \tag{3.8}$$

3.2.3 The Continuous Estimator, $G_3(t)$

Although F is discrete, the subnetwork's actual duration distribution is continuous. Therefore, it was anticipated that a continuous estimator might be in order. The estimator $G_3(t)$ is

continuous and equals $G_2(t)$ at every sampled point but interpolates linearly between sample points.

3.2.4 The Mixed Estimator, $G_4(t)$

The estimator $G_4(t)$ has the advantages of a continuous estimator between sampled values but preserves the discrete nature at the sampled values. Like $G_3(t)$, $G_4(t)$ also equals $G_2(t)$ at every sampled point. However, at each sampled point x_i, $G_4(t)$ has a jump of size p_i. For t between x_i and x_{i+1}, $G_4(t)$ interpolates linearly between $G_2(x_i)$ and $G_2(x_{i+1}) - p_{i+1}$.

3.2.5 The Mixed Estimator, $G_5(t)$

Like $G_4(t)$, the estimator $G_5(t)$ also assigns the discrete jump sizes to the sampled points. However, this estimator spreads the probability that is unaccounted for by the sampled values evenly over the range x_1 to x_m.

3.3 Criteria for a Good Estimator

The quantity being estimated is a distribution function. Thus, a "good" estimator, say G(t), should have the properties of a distribution function; namely,

(1) $0 \leq G(t) \leq 1$, $-\infty < t < \infty$ and

(2) $G(t_i) \leq G(t_j)$ for $t_i < t_j$.

In addition, it is desirable (but not requisite) that the estimator be "consistent" in the sense that the estimate of the distribution

function based on a sample containing all possible x_i is the true distribution function of X.

All of the estimators exhibit the properties of a distribution function if the sampling is performed without replacement. However, the estimators $G_4(t)$ and $G_5(t)$ may not be between zero and one for all t if sampling is with replacement.

Only estimators $G_2(t)$, $G_4(t)$, and $G_5(t)$ are always consistent. This follows immediately since if m = M and the sample corresponds to all M activity duration configurations, then

$$\sum_{i=1}^{m} p_i = 1 \ . \tag{3.9}$$

Furthermore only $G_2(t)$ satisfies

$$\lim_{m \to \infty} G(t) = F(t) \tag{3.10}$$

if the sampling is done with replacement.

Since the mean and upper percentiles of the subnetwork's duration distribution are the quantities of primary interest, this work also sought an estimator whose estimates of a distribution function's mean, μ, 90-th percentile, P_{90}, and 95th percentile, P_{95}, exhibit a high degree of precision. The simulation study described in subsection 3.5 was designed to determine the suitability of the proposed estimators in this regard.

3.4 Choosing Between Simple Random or Systematic Sampling

On the basis of computational ease alone, sampling with replacement is the preferred method. However, only $G_2(t)$ satisfies (3.10) if the sampling is done with replacement. Also $G_4(t)$ and $G_5(t)$ are not necessarily distribution functions if the sampling is done with replacement. On the other hand, if systematic sampling is employed these difficulties do not arise. In addition, the simulation study described in subsection 3.5 indicates that estimates derived from systematic samples contain more information than estimates based on simple random sampling. This was anticipated since Cochran (1946) showed that for at least partially ordered populations the variance of $\bar{x} = \sum_{i=1}^{m} x_i/m$ under systematic sampling is always less than it is under simple random sampling. Hence, the algorithm does its sampling via the systematic technique if possible. Unfortunately, the way a computer represents integers in its memory makes systematic sampling impractical of $M > M_o = 2^{\alpha-1} - 1$ where α is the number of binary bits in an integer word for the particular machine being used. For most modern IBM computers, $\alpha = 32$, and thus $M_o = 2{,}147{,}483{,}647$. When $M > M_o$, the algorithm uses random sampling with replacement.

3.4.1 An Ordering Scheme

Unfortunately, the relative magnitude of the subnetwork duration corresponding to a particular activity duration configuration cannot be determined in general unless all configurations are considered

explicitly. Hence, the following approximate ordering scheme was devised.

Let $v = 1, \ldots, n_c$ index the 2^{n_c} configurations of activity durations corresponding to

(1) the duration of each activity not in the cluster being equal to either its upper duration or its lower duration depending on whether a lower bound or an upper bound, respectively, is being determined and

(2) the durations for the activities in the cluster being at each of the 2^{n_c} possible combinations of their upper and lower points.

The activity duration configuration whose corresponding subnetwork duration, say x_v, is approximately the v-th smallest subnetwork duration can be determined from $g_{n_c}(v)$ defined by

$g_{n_c}(v)$ = the k-th smallest binary integer containing exactly i "1" s

where i is the smallest integer such that

$$v \leq \sum_{j=1}^{i} \binom{n_c}{j} \tag{3.11}$$

and

$$k = v - \sum_{j=0}^{i-1} \binom{n_c}{j} \tag{3.12}$$

with

$$\binom{n_c}{j} = n_c!/(n_c - j)!j! \tag{3.13}$$

and

$$\sum_{j=0}^{-1} \binom{n_c}{j} \equiv 0 \ . \tag{3.14}$$

In particular the activity duration configuration corresponding to approximately the v-th smallest subnetwork duration has the j-th activity in the cluster equal to its lower duration if the j-th digit (counting from the least significant digit) in $g_{n_c}(v)$ is 0 and equal to its upper duration if the j-th digit in $g_{n_c}(v)$ is 1.

For $v = 1$, $g_{n_c}(v) = 0_2$ (base 2), so under the approximate ordering x_1 equals the subnetwork's duration when every activity has its lower duration which in fact is the smallest possible x-value. Similarly, $x_{2^{n_c}}$ is the subnetwork's duration when every activity has its upper duration and is the largest possible x-value. For $1 \leq v_s < v_t \leq 2^{n_c}$, x_{v_s} is not necessarily less than or equal to x_{v_t}. However, for v_s very much smaller than v_t the activity configuration corresponding to $g_{n_c}(v_t)$ has more activities at their upper duration than the one corresponding to $g_{n_c}(v_s)$. Hence, x_{v_t} is likely to be larger than x_{v_s}. For example, Table 3 gives the approximate ordering of the x-values for the small subnetwork pictured in Figure 8 and described in Table 2 for the case when all five of the subnetwork's activiites are in the cluster C.

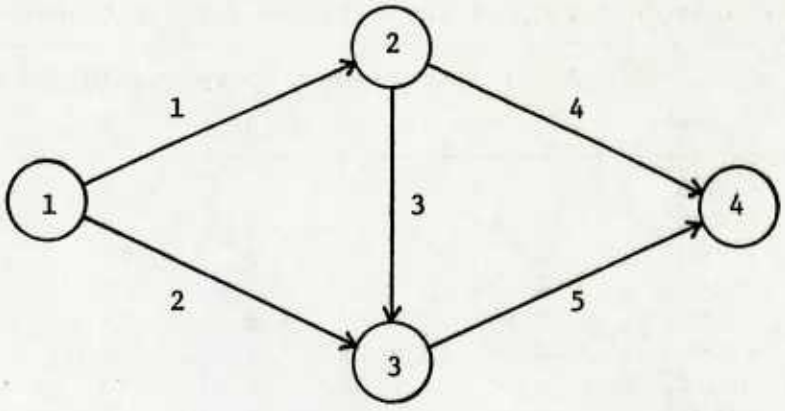


Figure 8

A small subnetwork.

TABLE 2

The Activity Durations for the Subnetwork in Figure 8

Activity	ℓ	u
1	5	7
2	8	10
3	3	6
4	5	6
5	4	8

TABLE 3

The Approximate Ordering of the x-values for the Subnetwork in Figure 8

		Activity durations corresponding to $g_5(v)$					
v	$g_5(v)$	5	4	3	2	1	x_v
1	00000_2	4	5	3	8	5	12
2	00001_2	4	5	3	8	7	14
3	00010_2	4	5	3	10	5	14
4	00100_2	4	5	6	8	5	15
5	01000_2	4	6	3	8	5	12
6	10000_2	8	5	3	8	5	16
7	00011_2	4	5	3	10	7	14
8	00101_2	4	5	6	8	7	17
9	00110_2	4	5	6	10	5	15
10	01001_2	4	6	3	8	7	14
11	01010_2	4	6	3	10	5	14
12	01100_2	4	6	6	8	5	15
13	10001_2	8	5	3	8	7	18
14	10010_2	8	5	3	10	5	18
15	10100_2	8	5	6	8	5	19
16	11000_2	8	6	3	8	5	16
17	00111_2	4	5	6	10	7	17
18	01011_2	4	6	3	10	7	14
19	01101_2	4	6	6	8	7	17
20	01110_2	4	6	6	10	5	15
21	10011_2	8	5	3	10	7	18
22	10101_2	8	5	6	8	7	21
23	10110_2	8	5	6	10	5	19
24	11001_2	8	6	3	8	7	18
25	11010_2	8	6	3	10	5	18
26	11100_2	8	6	6	8	5	19
27	01111_2	4	6	6	10	7	17
28	10111_2	8	5	6	10	7	21
29	11011_2	8	6	3	10	7	18
30	11101_2	8	6	6	8	7	21
31	11110_2	8	6	6	10	5	19
32	11111_2	8	6	6	10	7	21

3.4.2 Implementing the Ordering Scheme

An efficient algorithm for finding the k-th smallest binary integer containing exactly i "1"s was developed. The algorithm is as follows:

1. Let

   NP = the number of binary digits whose values are as yet undetermined,

   NI = the number of the NP remaining digits that are to be assigned the value "1", and

   J = the location of the digit whose value is currently being determined as counted from the right.

2. Set $NP = n_c$, $NI = i$, $J = n_c$, $B = \binom{n_c}{i}$, $R = B - k$.

3. If $NI < 1$, assign the value "0" to all remaining digits and stop. Otherwise, set

$$B = B \times NI/NP$$

$$NI = NI - 1$$

$$NP = NP - 1$$

$$RR = R - B.$$

   If $RR \leq 0$, go to 4. Otherwise, go to 5.

4. Assign the J-th right-most digit the value "1". Set $J = J - 1$. Go to 3.

5. Assign the J-th right-most digit the value "0". Set

$$J = J - 1$$

$$R = RR$$

$$B = B \quad (NP - NI)/NP$$

$$NP = NP - 1$$

$$RR = R - B \ .$$

If $RR \leq 0$, go to 4. Otherwise, do 5 again.

3.5 The Simulation Study

Since a subnetwork's duration distribution is the distribution of the maximum path length, most subnetwork duration distributions are skewed left. Nevertheless, the behavior of the proposed estimators $G_1(t)$, ..., $G_5(t)$ was determined for samples drawn from populations exhibiting a variety of distributional shapes. Since the beta distribution with probability density function (p.d.f.)

$$B_{\alpha,\beta}(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1 - t)^{\beta-1} \qquad \alpha, \beta > 0 \ , \quad 0 \leq t \leq 1 \qquad (3.15)$$

is a finite range distribution which can assume a wide variety of shapes, the subnetwork duration distributions corresponded to

$$B_{23,2}(t), \ B_{8,2}(t), \ B_{5,5}(t), \ B_{2,8}(t), \ \text{and} \ B_{2,23}(t)$$

in the simulation study. (These represent the shapes highly skewed left, skewed left, symmetrical, skewed right, and highly skewed right, respectively, as indicated in Figure 9.)

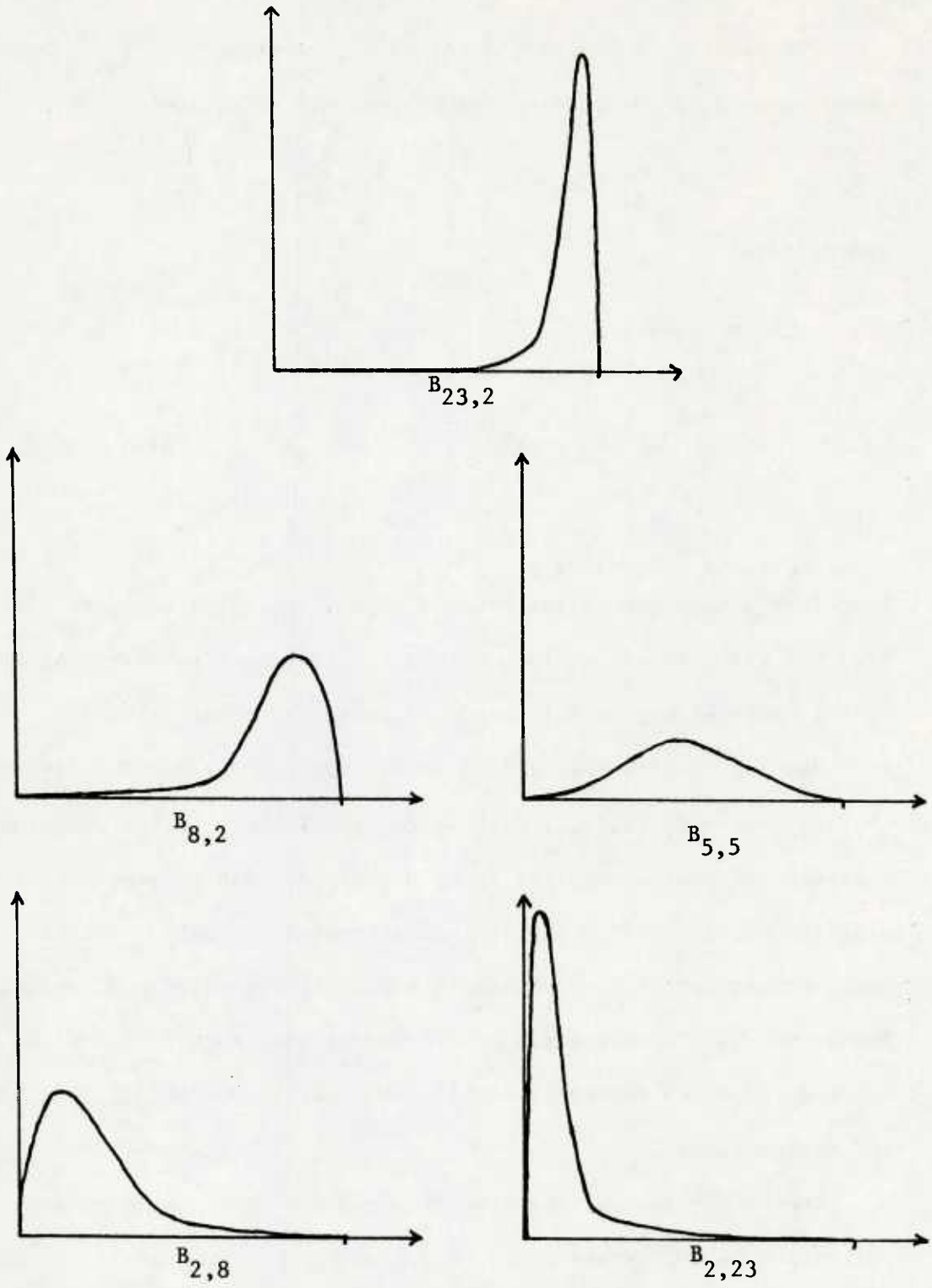


Figure 9

Duration distributions used in the simulation study.

The discrete subnetwork duration distribution, say $F_{\alpha,\beta}$, corresponding to $B_{\alpha,\beta}$ was constructed by randomly selecting

$$t_1 \leq t_2 \leq \cdots \leq t_{999} \leq t_{1000} = 1.0 ,$$

and defining

$$\begin{aligned} F_{\alpha,\beta}(t) &= 0 , && t < t_1 \\ &= \int_0^{t_i} B_{\alpha,\beta}(x)dx , && t_i \leq t < t_{i+1} \\ &= 1 , && t \geq 1.0 . \end{aligned} \qquad (3.16)$$

Therefore in the simulation study M = 1000. Also, since $\int_0^t B_{\alpha,\beta}(x)dx$ does not exist in closed form unless t = 1, it was evaluated at each $t_i < 1$ using an approximation due to Peizer and Pratt (1963).

Samples of size m were then taken from each of the $F_{\alpha,\beta}(t)$ and the estimators $G_1(t)$, ..., $G_5(t)$ were calculated. In the subnetwork analysis procedure, sampling is only employed when practical considerations dictate the explicit consideration of only a relatively small proportion of the subnetwork's activity duration configurations. Hence, the only sample sizes considered by this work were m = 10, 20, 50, and 100 which represent sampling proportions of 1%, 2%, 5%, and 10%, respectively.

The performance of the five proposed estimators with respect to estimation of the parameters μ, P_{95}, and P_{90} was evaluated for each of the $F_{\alpha,\beta}(t)$ on the basis of the following three criteria

(i) mean deviation,

$$MD(\hat{\gamma}) = \sum_{i=1}^{200} \frac{|\hat{\gamma}_i - \gamma|}{200}, \tag{3.17}$$

(ii) mean square error,

$$MSE(\hat{\gamma}) = \sum_{i=1}^{200} \frac{(\hat{\gamma}_i - \gamma)^2}{200}, \text{ and} \tag{3.18}$$

(iii) bias,

$$BIAS(\hat{\gamma}) = \sum_{i=1}^{200} \frac{\hat{\gamma}_i}{200} - \gamma \tag{3.19}$$

where in each case

γ = the parameter being estimated ($\gamma = \mu$, P_{95}, or P_{90});

$\hat{\gamma}_i$ = an estimate of γ based on the i-th sample of size m drawn; and

200 = the number of samples of size m drawn.

3.5.1 A Comparison of $G_1(t)$, ..., $G_5(t)$ under Systematic Sampling

Tables 4-8 indicate the simulation results for the proposed estimators under systematic sampling. The following observations may be made:

(1) For every case considered, $G_1(t)$ and $G_5(t)$ yielded highly biased estimates of at least one of the 3 parameters μ, P_{95}, and P_{90} while $G_2(t)$, $G_3(t)$, and $G_4(t)$ are only moderately biased.

TABLE 4

Simulation Results for the Highly Skewed Left $F_{23,2}(t)$ Having $\mu = 92122$, $P_{95} = 98763$, $P_{90} = 97805$*

Percent Sampling	Estimator	$\hat{\mu}$	$MD(\hat{\mu})$	$MSE(\hat{\mu})$	$BIAS(\hat{\mu})$	$\hat{P}_{95}$	$MD(\hat{P}_{95})$	$MSE(\hat{P}_{95})$	$BIAS(\hat{P}_{95})$	$\hat{P}_{90}$	$MD(\hat{P}_{90})$	$MSE(\hat{P}_{90})$	$BIAS(\hat{P}_{90})$
1%	G_1	50149	41972	17700	-41972	100000	1237	15	-1237	100000	2195	48	-2195
	G_2	91275	2592	99	-846	97518	2586	126	-1246	95593	3538	200	-2212
	G_3	85677	6454	508	-6445	95508	3471	201	-3255	93806	4241	269	-3999
	G_4	85724	6408	503	-6398	95575	3406	196	-3188	93888	4166	263	-3917
	G_5	50482	41640	17340	-41640	95200	3562	128	-3563	90523	7282	536	-7282
2%	G_1	49986	42135	17774	-42135	100000	1237	15	-1237	97642	1343	27	-164
	G_2	91480	1971	60	-641	97424	1801	56	-1339	96218	2189	90	-1587
	G_3	88844	3320	158	-3278	96381	2437	85	-2382	94870	2987	134	-2935
	G_4	88894	3286	155	-3227	96470	2350	81	-2293	94958	2903	130	-2847
	G_5	50928	41193	16972	-41193	95430	3334	114	-3334	91132	6674	458	-6674
5%	G_1	49884	42238	17843	-42238	97469	1294	21	-1294	92400	5405	296	-5405
	G_2	91902	1101	19	-220	98225	854	14	-539	97481	1182	22	-324
	G_3	90844	1524	33	-1277	97466	1335	29	-1297	96660	1375	30	-1145
	G_4	90895	1496	33	-1227	97531	1283	28	-1232	96735	1322	28	-1070
	G_5	52108	40014	16019	-40014	95897	2867	85	-2867	92388	5418	310	-5418
10%	G_1	49851	42271	17869	-42271	97010	1754	33	-1754	90963	6843	469	-6443
	G_2	91922	904	13	-200	98598	561	4	-165	97616	764	9	-189
	G_3	91375	1048	17	-746	98107	695	8	-656	97226	882	11	-579
	G_4	91429	1030	17	-692	98153	657	7	-609	97307	822	10	-498
	G_5	54153	37968	14442	-37968	96490	2274	55	-2274	93792	4012	179	-4012

* All entries have been multiplied by 10^5.

TABLE 5

Simulation Results for the Skewed Left $F_{8,2}(t)$ Having $\mu = 80127$, $P_{95} = 96008$, $P_{90} = 94100$*

Percent Sampling	Estimator	$\hat{\mu}$	$MD(\hat{\mu})$	$MSE(\hat{\mu})$	$BIAS(\hat{\mu})$	$\hat{P}_{95}$	$MD(\hat{P}_{95})$	$MSE(\hat{P}_{95})$	$BIAS(\hat{P}_{95})$	$\hat{P}_{90}$	$MD(\hat{P}_{90})$	$MSE(\hat{P}_{90})$	$BIAS(\hat{P}_{90})$
1%	G_1	50149	29978	9071	-29978	100000	3992	159	-1992	100000	5900	348	-5900
	G_2	79420	3674	203	-707	92007	5005	384	-4001	90558	5024	416	-3542
	G_3	73134	7247	711	-6993	89697	6356	563	-6310	87207	6949	666	-6893
	G_4	73184	7206	704	-6944	89770	6292	554	-6238	87270	6895	658	-6830
	G_5	50374	29754	8853	-29754	95034	981	10	-973	90127	3973	158	-3973
2%	G_1	49986	30141	9105	-30141	100000	3992	159	-3992	97641	3542	152	-3542
	G_2	79937	2305	101	-191	93913	3182	179	-2095	92167	3204	184	-1933
	G_3	77170	3617	190	-2957	92033	4031	255	-3975	90077	4155	280	-4022
	G_4	77220	3584	187	-2907	92097	3984	251	-3911	90133	4114	276	-3967
	G_5	50679	29449	8672	-29449	95066	959	9	-941	90261	3839	148	-3839
5%	G_1	49884	30243	9150	-30243	97469	1467	26	1461	92400	1699	32	-1699
	G_2	79795	1626	43	-332	95200	1762	47	-808	93135	1826	51	-964
	G_3	78726	2001	62	-1401	94201	2011	63	-1807	92188	2174	68	-1912
	G_4	78777	1974	61	-1350	94267	1971	61	-1741	92246	2134	66	-1854
	G_5	51543	28584	8172	-29594	05134	900	9	-874	90595	3505	125	-3505
10%	G_1	49851	30277	9168	-30277	97010	1002	12	1002	90963	3137	99	-3137
	G_2	79708	1987	53	-420	95166	1140	21	-842	93456	1668	41	-644
	G_3	79167	2157	62	-960	94745	1382	29	-1263	92850	1786	45	-1250
	G_4	79220	2143	61	-907	94822	1344	28	-1186	92902	1760	44	-1198
	G_5	53046	27081	7340	-27081	95224	820	8	-783	91076	3023	96	-3033

* All entries have been multiplied by 10^5.

TABLE 6

Simulation Results for the Symmetric $F_{5,5}(t)$ Having $\mu = 50107$, $P_{95} = 74892$, $P_{90} = 69918$*

Percent Sampling	Estimator	$\hat{\mu}$	$MD(\hat{\mu})$	$MSE(\hat{\mu})$	$BIAS(\hat{\mu})$	$\hat{P}_{95}$	$MD(\hat{P}_{95})$	$MSE(\hat{P}_{95})$	$BIAS(\hat{P}_{95})$	$\hat{P}_{90}$	$MD(\hat{P}_{90})$	$MSE(\hat{P}_{90})$	$BIAS(\hat{P}_{90})$
1%	G_1	50149	2502	84	42	100000	25108	6304	25108	100000	30082	9049	30082
	G_2	49798	3664	205	-309	71385	5230	645	-3507	67119	7398	898	-2799
	G_3	43308	6985	658	-6800	67175	8260	1023	-7718	62090	8498	1113	-7828
	G_4	43360	6939	651	-6748	67251	8192	1011	-7642	62160	8437	1100	-7758
	G_5	50133	34	0	25	94973	20080	4032	20080	89947	20029	4012	20029
2%	G_1	49986	1224	21	-121	100000	25108	6304	25108	97642	27724	7713	27724
	G_2	49808	3334	164	-299	73208	4345	259	-1684	68825	4068	237	-1093
	G_3	46864	4075	264	-3243	70871	4926	357	-4022	66021	4974	346	-3897
	G_4	46918	4050	261	-3189	70938	4876	351	-3955	66085	4934	341	-3833
	G_5	50130	60	0	22	94920	20028	4011	20027	89850	19932	3973	19932
5%	G_1	49884	514	3	-223	97469	22576	5101	22576	92400	22482	5058	22482
	G_2	49743	2328	76	-364	74020	2527	97	-872	69763	2683	115	-155
	G_3	48610	2530	97	-1497	72953	2923	133	-1939	68610	2665	115	-1308
	G_4	48665	2515	95	-1442	73019	2873	118	-1873	68673	2632	113	-1245
	G_5	50122	117	0	14	94761	19869	3948	19869	89600	19682	3874	19682
10%	G_1	49851	307	1	257	97010	22117	4894	22117	90963	21044	4430	21044
	G_2	50026	1251	21	-81	74652	1564	41	-240	69884	2006	64	-34
	G_3	49446	1286	25	-661	74138	1782	48	-754	69281	1999	63	-637
	G_4	49503	1278	24	-603	74194	1760	47	-699	69343	1975	62	-575
	G_5	50136	123	0	28	94570	19677	3872	19677	89288	19370	3753	19370

* All entries have been multiplied by 10^5.

TABLE 7

Simulation Results for the Skewed Right $F_{2,8}(t)$ Having $\mu = 20067$, $P_{95} = 42929$, $P_{90} = 36912$*

Percent Sampling	Estimator	$\hat{\mu}$	$MD(\hat{\mu})$	$MSE(\hat{\mu})$	$BIAS(\hat{\mu})$	$\hat{P}_{95}$	$MD(\hat{P}_{95})$	$MSE(\hat{P}_{95})$	$BIAS(\hat{P}_{95})$	$\hat{P}_{90}$	$MD(\hat{P}_{90})$	$MSE(\hat{P}_{90})$	$BIAS(\hat{P}_{90})$
1%	G_1	50149	30082	9133	3082	100000	57071	32571	57071	100000	63088	39801	63088
	G_2	20209	3583	181	141	41767	5436	490	-1162	35497	6699	718	-1414
	G_3	15120	5244	389	-4946	37129	6688	671	-5800	30775	7338	871	-6136
	G_4	15148	5220	385	-4919	37188	6640	663	-5741	30839	7285	861	-6073
	G_5	49894	29827	8897	29827	94794	52044	27086	52044	89948	53036	28128	53036
2%	G_1	49986	29919	8972	29919	100000	57071	32571	57071	97641	60730	36908	60730
	G_2	19867	2283	84	-200	41938	3809	221	-991	36214	4003	268	-698
	G_3	17288	3233	156	-2779	39630	4479	304	-3299	33979	4576	314	-2933
	G_4	17332	3200	153	-2735	39683	4446	301	-3247	34041	4534	309	-2870
	G_5	49562	29495	8700	29495	94918	51988	27028	51988	89841	52929	28015	52929
5%	G_1	49884	29817	8894	29817	97469	54540	29750	54540	92400	55489	30793	55489
	G_2	20034	1375	32	-33	42323	2114	72	-606	36311	2596	102	-600
	G_3	19008	1793	42	-1059	41319	2390	87	-1610	35474	2739	113	-1437
	G_4	19058	1764	41	-1008	41368	2371	84	-1561	35530	2707	111	-1382
	G_5	48635	28568	8162	28568	94753	51823	26857	51823	89594	52682	27755	52682
10%	G_1	49851	29784	8871	29784	97010	54080	29249	54080	90963	54051	29216	54051
	G_2	20093	960	13	26	42990	1788	48	61	36832	1778	48	-380
	G_3	19563	1062	15	-504	42488	1780	47	-441	36122	1902	54	-790
	G_4	19615	1040	15	-452	42538	1781	47	-391	36183	1867	52	-729
	G_5	47141	27075	7332	27075	94547	51617	26643	51617	89182	52270	27322	52270

* All entries have been multiplied by 10^5.

TABLE 8

Simulation Results for the Highly Skewed Right $F_{2,23}(t)$ Having $\mu = 8072$, $P_{95} = 18309$, $P_{90} = 15444$*

Percent Sampling	Estimator	$\hat{\mu}$	$MD(\hat{\mu})$	$MSE(\hat{\mu})$	$BIAS(\hat{\mu})$	$\hat{P}_{95}$	$MD(\hat{P}_{95})$	$MSE(\hat{P}_{95})$	$BIAS(\hat{P}_{95})$	$\hat{P}_{90}$	$MD(\hat{P}_{90})$	$MSE(\hat{P}_{90})$	$BIAS(\hat{P}_{90})$
1%	G_1	50149	42077	17788	42077	100000	81691	66734	81691	100000	84556	71497	84556
	G_2	6543	2188	74	-1529	15888	4864	369	-242	13125	5164	399	-2319
	G_3	3950	4139	192	-4123	12584	5965	519	-5725	9742	5981	499	-5702
	G_4	3894	4192	197	-4178	12648	5909	511	-5662	9795	5934	493	-5649
	G_5	49724	41652	17350	41652	94966	76657	58762	76657	89932	74488	55484	74488
2%	G_1	49986	41914	17588	41914	100000	81691	66734	81691	97642	82197	67591	82197
	G_2	7664	1217	23	-408	17050	3421	173	-126	14211	2867	126	-1233
	G_3	5749	2404	73	-2322	14976	3706	206	-2224	12394	3532	173	-3050
	G_4	5757	2394	72	-2315	15037	3656	201	-3272	12453	3489	168	-2991
	G_5	49257	41184	16963	41184	94908	76599	58674	76599	89825	74381	55325	74381
5%	G_1	49884	41812	17486	41812	97469	79160	62667	79160	92400	76956	59226	76956
	G_2	7964	858	12	-108	17740	2256	68	-569	15147	1821	47	-297
	G_3	7090	1232	22	-983	16952	2238	67	-1357	14276	1934	52	-1169
	G_4	7123	1203	21	-949	17011	2203	65	-1298	14330	1906	50	-1114
	G_5	47970	39897	15922	39897	94746	76437	58426	76437	89587	74141	54970	74141
10%	G_1	49851	41779	17544	41779	97010	78700	61940	78700	90963	75518	57031	75518
	G_2	7953	579	6	-119	17991	1054	18	-318	15410	1085	16	-34
	G_3	7469	762	10	-603	17476	1221	22	-833	14854	1059	16	-590
	G_4	7513	738	9	-559	17527	1190	21	-782	14896	1038	16	-548
	G_5	45895	37823	14311	37823	94548	76238	58123	76238	89198	73754	54398	73754

* All entries have been multiplied by 10^5.

(2) In the vast majority of the cases considered, $G_3(t)$ and $G_4(t)$ performed virtually the same in all respects while $G_2(t)$ performed approximately twice as well as either $G_3(t)$ or $G_4(t)$.

(3) The relative performances of $G_1(t)$, ..., $G_5(t)$ remained virtually unchanged as m increased.

(4) As would be expected, all five estimators increased in precision as sample size increased.

On the basis of these observations, $G_2(t)$ appears to be the "best" estimator and is the one implemented by the subnetwork analysis procedure.

3.5.2 The Performance of $G_2(t)$ under Systematic and Ramdom Sampling

Table 9 presents a comparison of the performance of the estimator $G_2(t)$ for both systematic sampling and random sampling under a variety of sampling conditions. In almost every case, systematic sampling was superior to random sampling and hence is the preferred technique.

TABLE 9

Ratios of the Empirical Behavior of $G_2(t)$ Using Systematic Sampling to that Using Random Sampling

Percent Sampling	Distribution Sampled	$MD(\hat{\mu})$	$MSE(\hat{\mu})$	$VAR(\hat{\mu})$	$BIAS(\hat{\mu})$	$MD(\hat{P}_{95})$	$MSE(\hat{P}_{95})$	$VAR(\hat{P}_{95})$	$BIAS(\hat{P}_{95})$	$MD(\hat{P}_{90})$	$MSE(\hat{P}_{90})$	$VAR(\hat{P}_{90})$	$BIAS(\hat{P}_{90})$
1%	$F_{8,2}$	.5065	.2416	.2581	.2613	.5860	.2748	.2660	.5370	.5646	.2591	.2819	.4677
	$F_{5,5}$	.5472	.2804	.2813	.4256	.5506	.2986	.3425	.4397	.6902	.4827	.5582	.4472
	$F_{2,8}$	.6163	.3514	.3535	.2655	.5379	.3037	.3210	.3265	.6721	.4680	.4830	.4722
2%	$F_{8,2}$	.4958	.2664	.2724	.1744	.6764	.3800	.4096	.5607	.6690	.3602	.3772	.5559
	$F_{5,5}$	.7842	.5655	.5620	4.4626	.6850	.3984	.4408	.4737	.6317	.3405	.3668	.3863
	$F_{2,8}$	.5858	.3373	.3374	-.4000	.5599	.3031	.3063	.5132	.5955	.3681	.3651	.9369
5%	$F_{8,2}$	.5932	.3440	.3378	.8058	.6447	.3730	.4166	.4635	.6010	.3187	.3387	.5020
	$F_{5,5}$	.9260	.7835	.7732	-15.1666	.7586	.4663	.4545	.8635	.7185	.5324	.5476	.2069
	$F_{2,8}$	.6255	.4102	.4102	1.2692	.5039	.2696	.2635	.6615	.6204	.3541	.3438	.1047
10%	$F_{8,2}$	1.1572	1.1521	1.1555	1.6627	1.5619	2.2203	2.8222	5.8276	5.7114	.3281	.2500	.9418
	$F_{5,5}$	.7099	.4468	.4468	1.3728	.5955	.3628	.3703	.3283	.6487	.4571	.4812	.4057
	$F_{2,8}$	.6387	.3513	.3611	.8929	.6441	.4247	.4247	.9838	.6515	.4000	.3916	-6.7857

4. ESTIMATION OF A DISCRETE DISTRIBUTION FUNCTION BY EXTRAPOLATING UPPER AND LOWER BOUNDS

4.1 Introduction

Since it is sometimes impractical to completely enumerate a subnetwork's discrete duration distribution, F, the bounds $F^{+}(t; \theta, \lambda)$ and $F^{-}(t; \theta, \lambda)$ are calculated as a first step in the determination of an estimate, $\hat{F}$, of F. Theorems 4 and 5 of section 2 imply that for θ very large both $F^{+}(t; \theta, \lambda)$ and $F^{-}(t; \theta, \lambda)$ may serve as adequate estimates of F. Unfortunately, it becomes increasingly laborious to calculate these quantities as $\theta \to \infty$. Hence the extrapolation procedure described in this section was devised as a practical alternative to evaluating the upper and lower bounds for large θ.

4.2 The Extrapolation Problem

Suppose that for a particular subnetwork, the numerical values of $F^{+}(t; \theta, \lambda)$ and $F^{-}(t; \theta, \lambda)$ are available for each of the combinations of $t = t_1, \ldots, t_I$ and $(\theta, \lambda) = (\theta_1, \lambda_1), \ldots, (\theta_J, \lambda_J)$ where

(1) $t_i \leq t_{i+1}$ for all i and

(2) $\theta_j \leq \theta_{j+1}$ and $\lambda_j \leq \lambda_{j+1}$ for all j.

The specific goal of the extrapolation procedure is to estimate F at the points $t_1, t_2, \ldots, t_I$.

Let

$$\omega = 1/(1 + \theta)$$

$$\omega_j = 1/(1 + \theta_j) , \qquad j = 1, \ldots, J , \tag{4.1}$$

and define

$$\text{(a) } H^{+}(t, \omega) = F^{+}(t; \theta, \lambda) \quad \text{and}$$

$$\text{(b) } H^{-}(t, \omega) = F^{-}(t; \theta, \lambda) \ . \tag{4.2}$$

Then the results of Theorems 4 and 5 can be restated as follows

(a) $H^{+}(t, \omega)$ is a nondecreasing function of ω for every t; $H^{-}(t, \omega)$ is a nonincreasing function of ω for every t;

(b) for any ω and t

$$H^{+}(t, \omega) \geq F(t) \geq H^{-}(t, \omega) \ ;$$

and

(c) there exists a finite value ω^* such that $\omega \leq \omega^*$ implies $H^{+}(t, \omega) = H^{-}(t, \omega) = F(t)$ for every t.

Thus, estimating $H^{+}(t, 0)$ and $H^{-}(t, 0)$ is the same as estimating $F(t)$. Although viable estimates of $F(t)$ can be obtained in a variety of ways, the proposed procedure uses the known quantities $H^{+}(t_i, \omega_j)$ and $H^{-}(t_i, \omega_j)$ $(i = 1, \ldots, I;\ j = 1, \ldots, J)$ to estimate functions $\hat{H}^{+}(t, \omega)$ and $\hat{H}^{-}(t, \omega)$ satisfying

(1) $\hat{H}^{+}(t_i, \omega_j) \geq \hat{H}^{+}(t_i, \omega_{j+1})$ for each i and j;

(2) $\hat{H}^{-}(t_i, \omega_j) \leq \hat{H}^{-}(t_i, \omega_{j+1})$ for each i and j; and

(3) $\hat{H}^{+}(t_i, 0) = \hat{H}^{-}(t_i, 0) \leq \hat{H}^{+}(t_{i+1}, 0) = \hat{H}^{-}(t_{i+1}, 0)$ for each i. (4.3)

and then estimates $F(t_i)$ by

$$\hat{F}(t_i) = \hat{H}^+(t_i, 0) = \hat{H}^-(t_i, 0) \qquad i = 1, \ldots, I . \tag{4.4}$$

The basic idea is simply for each t_i to fit a function $\hat{H}^+(t_i, \omega)$ (as a function of ω) to the sequence of upper bounds on $F(t_i)$ namely $H^+(t_i, \omega_1), \ldots, H^+(t_i, \omega_J)$ and also fit a function $\hat{H}^-(t_i, \omega)$ to the lower bounds on $F(t_i)$ under the restriction that

$$\lim_{\omega \to 0} \hat{H}^+(t_i, \omega) = \lim_{\omega \to 0} \hat{H}^-(t_i, \omega) . \tag{4.5}$$

Since $F(t_1) \leq \ldots \leq F(t_I)$, the additional restriction that

$$\hat{H}^+(t_1, 0) \leq \ldots \leq \hat{H}^+(t_I, 0) \tag{4.6}$$

is imposed so that

$$\hat{F}(t_1) \leq \ldots \leq \hat{F}(t_I) . \tag{4.7}$$

4.3 A Linear Programming Solution to the Extrapolation Problem

The determination of $\hat{H}^+$ and $\hat{H}^-$ is as follows. For each i let

$$\hat{H}^+(t_i, \omega) = \alpha_{0i} + \alpha_{1i}\omega + \alpha_{2i}\omega^2 \tag{4.8}$$

and

$$\hat{H}^-(t_i, \omega) = \beta_{0i} + \beta_{1i}\omega + \beta_{2i}\omega^2 \tag{4.9}$$

where α_{0i}, α_{1i}, α_{2i}, β_{0i}, β_{1i}, and β_{2i} $i = 1, \ldots, I$ are all constants determined so that (4.3) holds. Since (4.4) is a quadratic function in ω, requirement (1) of (4.3) is met by requiring

$$\alpha_{1i} \geq 0 \qquad i = 1, \ldots, I \ . \tag{4.10}$$

Similarly, requirement (2) of (4.3) is met by restricting

$$\beta_{1i} \leq 0 \qquad i = 1, \ldots, I \ . \tag{4.11}$$

Finally, requirement (3) is met by requiring

$$\alpha_{0i} = \beta_{0i} \leq \alpha_{0,i+1} = \beta_{0,i+1} \qquad i = 1, \ldots, I \ . \tag{4.12}$$

Of course, when the restrictions (4.10), (4.11), and (4.12) are enforced, it is not always possible to have

$$\begin{aligned} &(1)\ \hat{H}^{+}(t_i, \omega_j) = H^{+}(t_i, \omega_j) \text{ and} \\ &(2)\ \hat{H}^{-}(t_i, \omega_j) = H^{-}(t_i, \omega_j) \end{aligned} \tag{4.13}$$

for all i and j. Hence, the constants α_{0i}, α_{1i}, α_{2i}, β_{0i}, β_{1i}, and β_{2i} (i = 1, ..., I) are determined by minimizing

$$\sum_{j=1}^{J} a(\omega_j)\Big[\sum_{i=1}^{I} \big(\big|\hat{H}^{+}(t_i, \omega_j) - H^{+}(t_i, \omega_j)\big| + \big|\hat{H}^{-}(t_i, \omega_j) - H^{-}(t_i, \omega_j)\big|\big)\Big] \tag{4.14}$$

under the restrictions (4.10) - (4.12) where the $a(\omega_j)$ is a specified nonnegative weighting constant.

The weights, $a(\omega)$, in (4.14) should reflect the increase in information about F(t) as $\omega \to 0$ (i.e., $\theta \to \infty$). In the algorithm the weight $a(\omega)$ has been defined to be

$$a(\omega) = 1 - 2\omega^2 + 3\omega^3 \ .$$

The coefficients in the cubic function were selected so

$$a(0) = 1,\ a(1) = 0,\ \text{and}\ \left.\frac{da(\omega)}{d\omega}\right|_{\omega=0} = \left.\frac{da(\omega)}{d\omega}\right|_{\omega=1} = 0 \tag{4.15}$$

Hence, the points $\omega = .25$ and $\omega = .5$ which correspond to $\theta = 3$ and $\theta = 1$, respectively, have weights .84375 and .50000, respectively.

The minimization of (4.14) subject to (4.10) - (4.12) can be restated as

$$\text{minimize} \quad \sum_{j=1}^{J} a(\omega_j) \sum_{i=1}^{I} (u_{ij} + v_{ij})$$

subject to

$$-u_{ij} \leq \sum_{k=1}^{i} \gamma_k + \alpha_{1i}\omega_j + (\alpha_{2i1} - \alpha_{2i2})\omega_j^2 - H^{+}(t_i, \omega_j) \leq u_{ij} \quad \text{for all } i, j$$

$$-v_{ij} \leq \sum_{k=1}^{i} \gamma_k - \beta_{1i}\omega_j + (\beta_{2i1} - \beta_{2i2})\omega_j^2 - H^{-}(t_i, \omega_j) \leq v_{ij} \quad \text{for all } i, j$$

$$\sum_{k=1}^{I} \gamma_k \leq 1 \tag{4.16}$$

$$u_{ij},\ v_{ij},\ \alpha_i,\ \beta_{1i},\ \alpha_{1i},\ \alpha_{2i1},\ \beta_{2i2},\ \beta_{2i1},\ \beta_{2i2} \geq 0 \quad \text{for all } i, j$$

where

$$\alpha_{2i} = \alpha_{2i1} - \alpha_{2i2}\ , \tag{4.17}$$

$$\beta_{2i} = \beta_{2i1} - \beta_{2i2}\ , \text{ and} \tag{4.18}$$

$$\alpha_{0i} = \beta_{0i} = \sum_{k=1}^{i} \gamma_k\ . \tag{4.19}$$

This is a linear programming problem which may be solved using any standard method. In the computer implementation of the subnetwork

analysis procedure, a streamlined version of the revised simplex algorithm was especially prepared and implemented to solve this problem. Once this linear programming problem has been solved, $\hat{F}$ is retrieved through the relation

$$\hat{F}(t_i) = \sum_{k=1}^{i} \gamma_k \quad \text{for all } i \ . \tag{4.20}$$

If the upper bounds $F^{+}(t, \theta, \lambda)$ and lower bounds $F^{-}(t; \theta, \lambda)$ were determined without sampling then

$$H^{+}(t_i, w_1) \geq \ldots \geq H^{+}(t_i, w_J) \geq F(t_i) \geq H^{-}(t_i, w_J) \geq \ldots \geq H^{-}(t_i, w_1) \ .$$

However this relationship does not necessarily hold if sampling is used in the determination of the bounds. Hence the determination of $\hat{H}^{+}(t, w)$ and $\hat{H}^{-}(t, w)$ does not include the restriction

$$\hat{H}^{-}(t_i, w_J) \leq \hat{F}(t_i) \leq \hat{H}^{+}(t_i, w_J) \qquad i = 1, \ldots, I \ .$$

It should also be noted that, if a weighted least squares criterion had been used instead of minimization of a weighted sum of absolute residuals, then the determination of $\hat{F}(t)$ would have been a quadratic programming problem instead of a somewhat simpler linear programming problem.

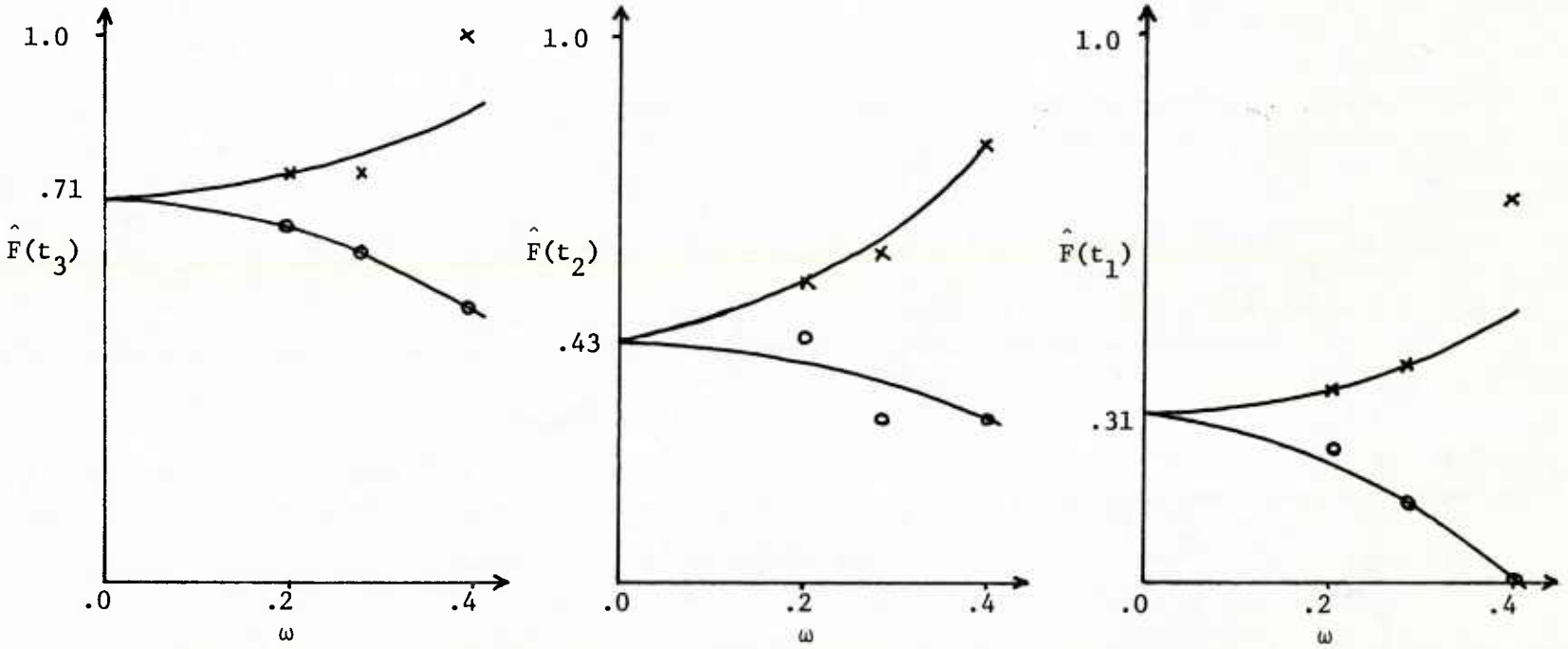
4.4 An Example of the Linear Programming Solution

Using the simplex algorithm referred to in subsection 4.3, the linear programming problem (4.16) was solved for the data in Table 10. Figure 10 indicates the fits obtained.

TABLE 10

Extrapolation Data

	t	$F^+(H^+)$	$F^-(H^-)$
$\theta_1 = 1.5$ ($\omega_1 = .4$)	t_1	.7	0.0
	t_2	.8	.3
	t_3	1.0	.5
$\theta_2 = 2.5$ ($\omega_2 = .2851$)	t_1	.4	.15
	t_2	.6	.3
	t_3	.75	.6
$\theta_3 = 3.5$ ($\omega_3 = .2222$)	t_1	.35	.25
	t_2	.55	.45
	t_3	.75	.65



As required, $\hat{F}(t_1) = .31 \leq \hat{F}(t_2) = .43 \leq \hat{F}(t_3) = .71$.

Figure 10

Extrapolation results for the data in Table 10.

5. POTENTIAL MODIFICATIONS OF THE SUBNETWORK ANALYSIS PROCEDURE

5.1 Introduction

The objective of Subnetwork Analysis is to determine each subnetwork's duration distribution, say $F(t)$. When this step is begun, each activity has a specified duration distribution. Let

n = number of activities in the subnetwork,

X_i = the duration of activity i, and

$F_{X_i}(t)$ = the c.d.f. for activity i.

Also, let

m = the number of paths through the subnetwork, and

Y_j = the length of the j-th path through the subnetwork

$$= \sum_{i=1}^{n} \delta_{ij} X_i \tag{5.1}$$

where

δ_{ij} = 1 if activity i is on the j-th path

= 0 otherwise.

Let the maximum path length be

$$Y^* = \max_{1 \le j \le m} Y_j . \tag{5.2}$$

Then

$$F(t) = P(Y^* \le t) = \int_{-\infty}^{t} \cdots \int_{-\infty}^{t} dF_{Y_1, \ldots, Y_m}(t_1, \ldots, t_m) \tag{5.3}$$

where

$F_{Y_1, \ldots, Y_m}(t_1, \ldots, t_m)$ = the joint distribution of the m paths.

The activity distributions are assumed to be independent. Thus, the marginal distribution of Y_j, say $F_{Y_j}(t)$, is the convolution of the path's activity duration distributions; that is,

$$F_{Y_j}(t) = \int \cdots \int F_{X_{j_1}}\left(t - \sum_{i \neq j_1} \delta_{ij} X_i\right) \prod_{\substack{i \neq j_1 \\ i \ni \delta_{ij}=1}} dF_{X_i}(X_i) \tag{5.4}$$

where j_1 is the index of an activity on the path. Furthermore, if

$$F_{Y_j|X}(t)$$

denotes the conditional distribution of Y_j given a set, X, of activity duration values, then

$$F_{Y_j|X}(t) = \int \cdots \int F_{X_{j_1}}\left(t - \sum_{i \neq j_1} \delta_{ij} X_i\right) \prod_{\substack{i \neq j_1 \\ i \ni \delta_{ij}=1 \\ X_i \notin X}} dF_{X_i}(X_i) \tag{5.5}$$

which is the convolution of the path's activity durations not in X. If there is no activity that is in two or more of $Y_{j_1}, \ldots, Y_{j_k}$ (that is, these paths have no activities in common), then $Y_{j_1}, \ldots, Y_{j_k}$ are independent, and

$$F_{Y_{j_1}, \ldots, Y_{j_k}}(t_{j_1}, \ldots, t_{j_k}) = \prod_{i=1}^{k} F_{Y_{j_1}}(t_{j_i}) . \tag{5.6}$$

However, if

$$X = \{X_i | \text{activity } i \text{ is in more than one of } Y_1, \ldots, Y_m\}$$

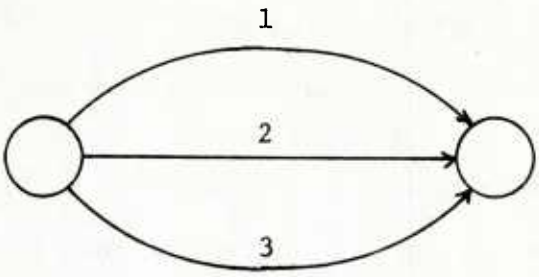
is a nonempty set, then $Y_1, \ldots, Y_m$ are dependent, and

$$F(t) = \int \cdots \int \left[\prod_{j=1}^{m} F_{Y_j|X}\left(t - \sum_{X_i \in X} \delta_{ij} X_i\right)\right] \prod_{X_i \in X} dF_{X_i}(X_i). \tag{5.7}$$

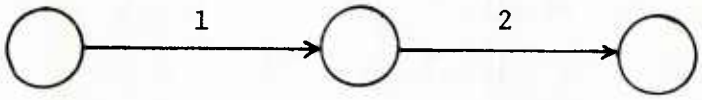
5.2 Explicit Evaluation of the Subnetwork Duration Distribution

For simple subnetworks it is relatively easy to identify all of the paths and give an explicit expression for $F(t)$ via (5.7). In particular Hartley and Wortham (1966) considered series, parallel, and Wheatstone Bridge subnetworks (see Figure 11). Ringer (1969) extended this work to include Double Wheatstone Bridge and Criss-Cross subnetworks (see Figure 12). These exact expressions for the subnetwork duration distribution form the basis of Step 2, Simplification, in the project scheduling procedure. Interestingly, this implies that the subnetworks actually considered in Step 4, Subnetwork Analysis, do not have any of these simple activity configurations in them, and hence are generally fairly complex.

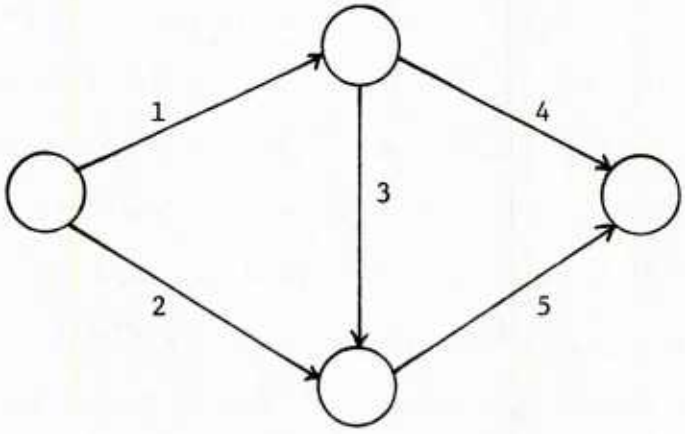
To utilize (5.7) to determine $F(t)$, all the paths through the subnetwork must be identified and then the numerical evaluation of (5.7) performed. Martin (1964) presented a clever method for performing the numerical evaluation of (5.7) when the activity duration distributions were all piecewise polynomial functions with finite ranges. Martin's technique is most readily suited to subnetworks primarily composed of activities in series or parallel. Unfortunately, the subnetworks generally encountered in the Subnetwork Analysis step are not of this form. Furthermore, Martin's technique becomes computationally impractical for large subnetworks.



Activities in Parallel



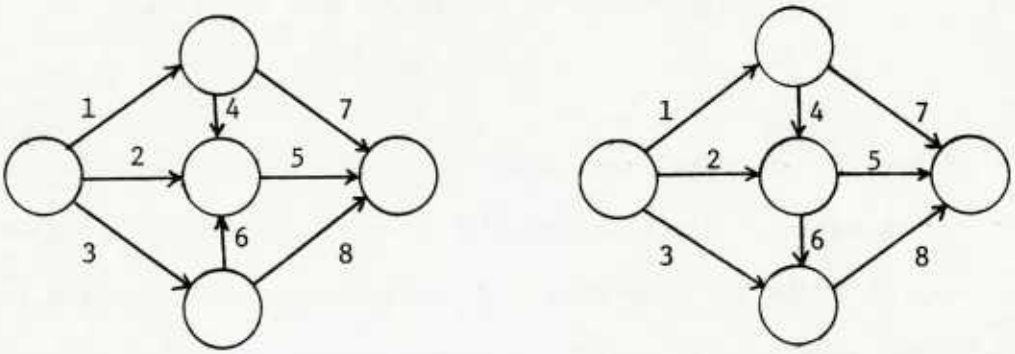
Two Activities in Series



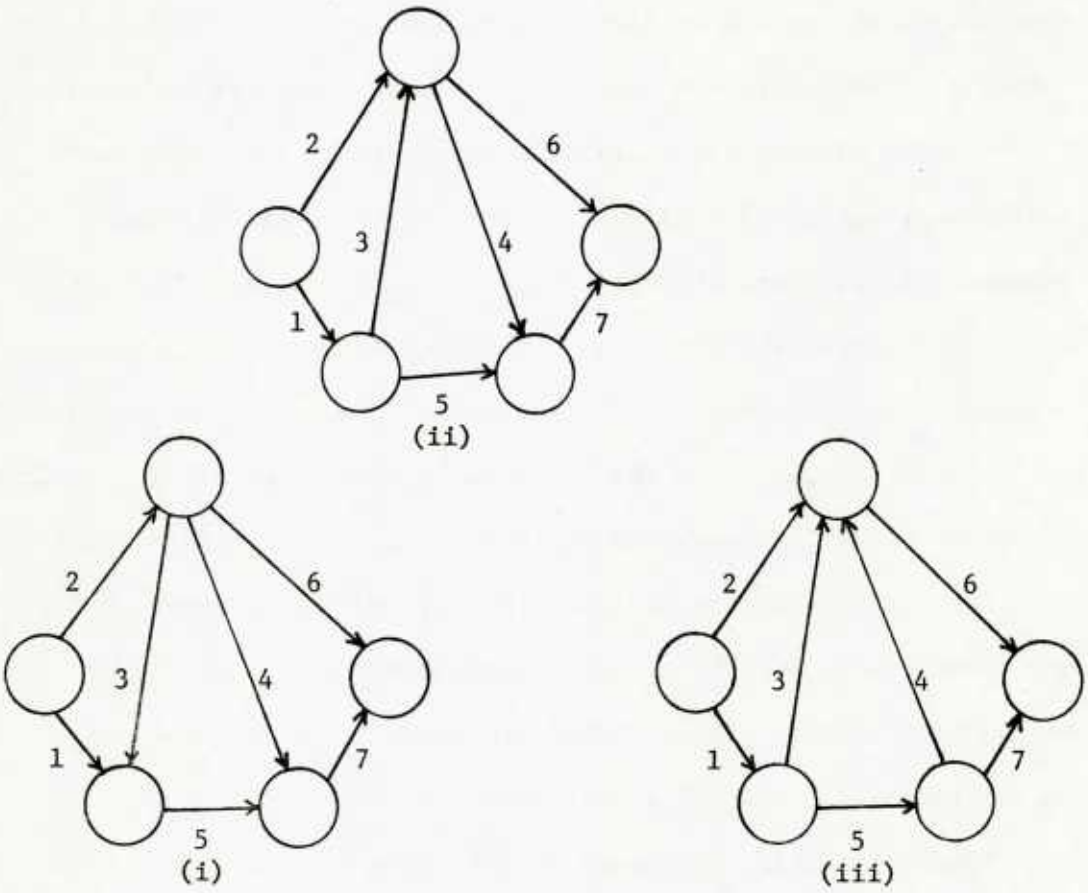
Wheatstone Bridge

Figure 11

Subnetworks considered by Hartley and Wortham (1966).



Double Wheatstone Bridge



Criss-Cross

Figure 12

Subnetworks considered by Ringer (1969).

5.3 Approximating the Subnetwork Duration Distribution $F(t)$

Since the explicit evaluation of the exact expression for the subnetwork duration distribution $F(t)$ given in (5.7) is generally impractical for other than simple subnetworks, several authors have considered approximating $F(t)$. A review of the classical approximation procedures is given in Moder and Phillips (1974). The more recent approximation procedures are essentially based on either sophisticated Monte Carlo simulation or the determination of upper and lower bounds for $F(t)$. The Subnetwork Analysis procedure developed in Sections 2 - 4 is one of these approximation procedures. That Subnetwork Analysis procedure basically estimates $F(t)$ by extrapolating a sequence of upper and lower bounds on the subnetwork's discrete duration distribution F - with specialized Monte Carlo techniques sometimes employed in the determination of the upper and lower bounds.

Noteworthy papers on the Monte Carlo simulation of $F(t)$ include Van Slyke (1963), Gaver and Burt (1968), and Burt and Garman (1971).

The two outstanding published techniques for determining upper and lower bounds on $F(t)$ are due to Robillard and Trahan (1977) and Kleindorfer (1971). These techniques are briefly discussed in subsections 5.3.1 and 5.3.2, respectively.

Subsection 5.3.3 indicates several ways that the Monte Carlo techniques and the upper and lower bounds of Kleindorfer (1971) and Robillard and Trahan (1977) can be incorporated into the general Subnetwork Analysis procedure.

5.3.1 Robillard and Trahan's Lower Bound on $F(t)$

Robillard and Trahan (1977) proposed a lower bound, $\bar{F}(t)$, for $F(t)$ based on a Bonferroni inequality. Specifically,

$$P(\max_{1\le j\le m} Y_j > t) = P(\bigcup_{j=1}^{m} \{Y_j > t\}) \le \sum_{j=1}^{m} P(Y_j > t), \tag{5.8}$$

so

$$\begin{aligned} F(t) = P(\max_{1\le j\le m} Y_j \le t) &= 1 - P(\max_{1\le j\le m} Y_j > t) \ge 1 - \sum_{j=1}^{m} P(Y_j > t) \\ &= 1 - \sum_{j=1}^{m} [1 - F_{Y_j}(t)] = 1 - m + \sum_{j=1}^{m} F_{Y_j}(t) \equiv \bar{F}(t). \end{aligned} \tag{5.9}$$

Robillard and Trahan (1977) evaluate the term $\sum_{j=1}^{m} F_{Y_j}(t)$ using the characteristic functions, say $\psi_{Y_1}(\tau)$, ..., $\psi_{Y_m}(\tau)$, of Y_1, ..., Y_m. Let I_t denote the integration corresponding to the inversion of a characteristic function. Then

$$\sum_{j=1}^{m} F_{Y_j}(t) = \sum_{j=1}^{m} I_t[\psi_{Y_j}] = I_t[\sum_{j=1}^{m} \psi_{Y_j}] \tag{5.10}$$

where the last equality follows from the linearity of integration. For example, if Y_1, ..., Y_m are all continuous random variables, then

$$\begin{aligned} \sum_{j=1}^{m} I_t[\psi_{Y_j}(\tau)] &= \sum_{j=1}^{m} \{\int_{-\infty}^{t}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-i\tau x}\psi_{Y_j}(\tau)d\tau dx\} \\ &= \int_{-\infty}^{t}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-i\tau x} \sum_{j=1}^{m} \psi_{Y_j}(\tau)d\tau dx \\ &= I_t[\sum_{j=1}^{m} \psi_{Y_j}(\tau)]. \end{aligned} \tag{5.11}$$

Although $\psi_{Y_j}(\tau)$ can be written as the product of the characteristic functions for the individual activities in Y_j, this approach to evaluating $\sum_{j=1}^{m} \psi_{Y_j}(\tau)$ would require the explicit enumeration of all the subnetwork's paths which is computationally impractical for large complex subnetworks. Therefore Robillard and Trahan (1977) developed the following recursive scheme for evaluating $\sum_{j=1}^{m} \psi_{Y_j}(\tau)$. Let $\psi_{X_1}(\tau), \ldots, \psi_{X_n}(\tau)$ denote the characteristic functions of the individual activity durations $X_1, \ldots, X_n$. Let the k-th activity originate at node $Orig_k$ and terminate at node $Term_k$. If

$$B_i = \{k | Term_k = i\}, \tag{5.12}$$

$$\phi(\tau,1) = 1, \text{ and} \tag{5.13}$$

$$\phi(\tau,i) = \sum_{k \epsilon B_i} \phi(\tau, Orig_k)\psi_{X_k}(\tau), \quad i = 2, \ldots, N, \tag{5.14}$$

where N is the number of nodes, then

$$\sum_{j=1}^{m} \psi_{Y_j}(\tau) = \phi(\tau,N) . \tag{5.15}$$

Although it is not explicitly noted by Robillard and Trahan (1977), the number of paths, m, can also be recursively generated. If

$$m_1 = 1 \tag{5.16}$$

and

$$m_i = \sum_{k \epsilon B_i} m_k \qquad i = 2, \ldots, N, \tag{5.17}$$

then

$$m = m_N. \tag{5.18}$$

Apart from any numerical inaccuracies in the computation (5.14), the tightness of the lower bound $\underline{F}(t)$ is the same as the tightness of the Bonferroni inequality (5.8).

Robillard and Trahan (1977) also note that another Bonferroni inequality implies that

$$P(\max_{1\le j\le m} Y_j > t) = P(\bigcup_{j=1}^{m} \{Y_j > t\}) \ge \sum_{j=1}^{m} P(Y_j > t) - \sum_{i<j} P(Y_i > t, Y_j > t) \tag{5.19}$$

Unfortunately, to use (5.19) as the basis for an upper bound on $\bar{F}(t)$ seems to require the explicit enumeration of the paths because of the joint nature of $P(Y_i > t, Y_j > t)$ and the lack of a convenient upper bound for $P(Y_i > t, Y_j > t)$.

5.3.2 Kleindorfer's Upper and Lower Bounds on $\bar{F}(t)$

Let the subnetwork's activities be numbered $i = 1, \ldots, n$ in such a way that, if $i < j$ and both activities i and j are on a path, then activity i precedes activity j. Let A_i denote the set of activities which immediately precede activity i on some path. As before, let

X_i = the duration of activity i

and also define

U_i = the earliest time at which activity i can commence,

$$P_i(t) = P(U_i \le t) \tag{5.20}$$

$V_i = U_i + X_i$ = the completion time for activity i, and

$Q_i = P(V_i \le t)$.

Kleindorfer (1971) proposed upper and lower bounds for $\bar{F}(t)$ by recursively defining upper bounds, $P_i'(t)$ and $Q_i'(t)$, and lower bounds, $P_i''(t)$ and $Q_i''(t)$ on $P_i(t)$ and $Q_i(t)$, respectively. The upper bounds $P_i'(t)$ are based upon the inequality

$$\min_{j \in A_i} P(V_j \leq t) \geq P(\max_{j \in A_i} V_j \leq t) = P_i(t), \tag{5.21}$$

and $Q_i'(t)$ is simply the convolution of $F_{X_i}(t)$ and $P_i'(t)$. The lower bounds $P_i''(t)$ are based upon the inequality

$$P_i(t) = P(\max_{j \in A_i} V_j \leq t) \geq \prod_{j \in A_i} P(V_j \leq t), \tag{5.22}$$

and $Q_i''(t)$ is the convolution of $F_{X_i}(t)$ and $P_i''(t)$. (Although Kleindorfer proves a version of (5.22), the inequality as stated follows from the more general results of Esary, Proschan, and Walkup (1967).)

The recursive relations for $P_i'(t)$, $Q_i'(t)$, $P_i''(t)$, and $Q_i''(t)$ are as follows: For notational convenience assume that activity 1 is an activity with zero duration which precedes the rest of the subnetwork and that activity n is an activity with zero duration which follows the completion of the rest of the subnetwork. Furthermore, assume that X_i is a discrete, nonnegative random variable taking on values in S for all i. Then, for $t \geq 0$,

$$P_1'(t) = P_1(t) = Q_1(t) = Q'(t) = 1, \tag{5.23}$$

$$P_i(t) = \min_{j \in A_i} Q_j'(t), \quad i = 2, \ldots, n, \tag{5.24}$$

$$Q_i'(t) = \sum_{s \in S} P(X_i = s) P_i'(t - s), \quad i = 2, \ldots, n, \tag{5.25}$$

$$P_1''(t) = P_1(t) = Q_1(t) = Q''(t) = 1, \tag{5.26}$$

$$P_i''(t) = \prod_{j \in A_i} Q_j''(t), \quad i = 2, \ldots, n, \text{ and} \tag{5.27}$$

$$Q_j''(t) = \sum_{s \in S} P(X_i = s) P_i''(t - s), \quad i = 2, \ldots, n. \tag{5.28}$$

Finally,

$$P''_n(t) \leq F(t) \leq P'_n(t). \tag{5.29}$$

The computational beauty of these bounds on $F(t)$ is that they do not necessitate the enumeration of all of the subnetwork's paths.

The tightness of these bounds on $F(t)$ depends on the structure of the subnetwork. Since the recursive relations (5.23) - (5.28) sequentially bound the $P_i(t)$ in terms of the bounds for the completion time distributions of the activities immediately preceding activity i, the differences $P'_i(t) - P_i(t)$ and $P_i(t) - P''_i(t)$ essentially cumulate as i increases. Therefore, the bounds on $F(t)$ will generally tend to be tighter the shorter the subnetwork's paths. Furthermore, the difference

$$\min_{j \in A_i} P(V_j \leq t) - P(\max_{j \in A_i} V_j \leq t) \tag{5.30}$$

tends to decrease as the V_j's have more and more activities in common; whereas, the difference

$$P(\max_{j \in A_i} V_j \leq t) - \prod_{j \in A_i} P(V_j \leq t) \tag{5.31}$$

tends to increase as the V_j's have more and more activities in common. Thus subnetwork structures that lead to tight upper bounds on $F(t)$, lead to loose lower bounds on $F(t)$, and vice versa. Of course, the tightness of both the upper and lower bounds tends to decrease as the number of paths increases.

5.3.3 Incorporating Different Methods of Approximating $F(t)$ into Subnetwork Analysis

The Monte Carlo simulation techniques and the bounding procedures of Kleindorfer (1971) and Robillard and Trahan (1977) referred to thus far in subsection 5.3 could be used to modify the current Subnetwork Analysis procedure discussed in Sections 2 - 4. Since the empirical experience with the modifications to be briefly described in the remainder of this subsection is generally extremely limited, these potential modifications are really subjects for future research.

A Monte Carlo simulation of the subnetwork duration distribution $F(t)$ could, of course, essentially replace the current Subnetwork Analysis procedure. A less radical revision would be to carry out the cluster formation procedure described in subsection 2.2 for a fixed (presumably large) value of (θ,λ); let IMPORTANT be the set of all activities in the union of the clusters; and then estimate $F(t)$ by fixing the durations of the activities not in IMPORTANT at their mean values and doing a Monte Carlo simulation of the durations for the activities in IMPORTANT. The durations of the activities in IMPORTANT could be simulated from either their actual distributions or their approximate two-point discrete distributions. Another potential modification would be to perform the current Subnetwork Analysis procedure as is except that the upper and lower bounds $F^{+}(C;t)$ and $F^{-}(C;t)$ used in determining $F^{-}(t;\theta,\lambda)$ and $F^{+}(t;\theta,\lambda)$ could be determined with the durations for activities in C determined by a Monte Carlo simulation of their actual duration distributions or to their two-point discrete distributions.

Another possible replacement for the current Subnetwork Analysis procedure would be to determine Kleindorfer's upper bound on $F(t)$ and either Kleindorfer's or Robillard and Trahan's lower bound on $F(t)$ and then use the average of these two bounding distributions as the estimate of $F(t)$. (Of course, the maximum of Kleindorfer's and Robillard and Trahan's lower bounds is also a valid lower bound.) Again a less radical revision would be to carry out the cluster formation procedure described in subsection 2.2 for a fixed (presumably large) value of (θ,λ); let IMPORTANT be the set of all activities in the union of the clusters; and then estimate $F(t)$ by fixing the durations of the activities not in IMPORTANT at their mean values and averaging the upper and lower bounds for the subnetwork duration distribution when the durations for the activities in IMPORTANT have either their actual distributions or their two-point discrete distributions. The durations for the activities not in IMPORTANT could, alternatively, be fixed at their lower values when the upper bound is being determined and be fixed at their upper values when the lower bound is being determined. Finally, another potential modification would be to perform the current Subnetwork Analysis procedure as is except that the upper and lower bounds $F^{+}(C;t)$ and $F^{-}(C;t)$ could be either Kleindorfer's or Robillard and Trahan's bounds determined with the durations for the activities in C having either their actual distributions or their two-point discrete distributions.

Presumably, a project scheduler might settle for a project schedule which has the probability of the project's completion by the specified deadline bounded from below by a specified amount.

In such instances lower bounds on the subnetwork duration distributions suffice. Then the Subnetwork Analysis procedure could be replaced by a procedure which simply determines either Kleindorfer's or Robillard and Trahan's lower bound. Alternatively, the cluster formation procedure could be carried out for a specified value of (θ,λ), the set IMPORTANT of all activities in the union of the clusters formed, and then either Kleindorfer's or Robillard and Trahan's lower bound computed with the durations for the activities outside IMPORTANT fixed and the durations for the activities in IMPORTANT having either their actual distributions or their two-point discrete distributions.

5.4 Additional Probability Inequalities as Bases for Upper and Lower Bounds on $\bar{F}(t)$

In addition to the ones cited in subsections 5.3.1 and 5.3.2, there are other known probability inequalities which imply upper and lower bounds on $\bar{F}(t) = P(\max_{1\le j\le m} Y_j \le t)$. Three upper bounds on $P(\max_{1\le j\le m} Y_j \le t)$ and the authors who proposed them are:

(i) Chung and Erdos (1952),

$$P(\max_{1\le j\le m} Y_j \le t) \le 1 - \{[\sum_{j=1}^{m} P(Y_j > t)]^2/[\sum_{j=1}^{m} P(Y_j > t) + \sum_{i\ne j} P(Y_i > t, Y_j > t)]\}; \tag{5.32}$$

(ii) Dawson and Sankoff (1967),

$$p(\max_{1\le j\le m} Y_j \le t) \le 1 - \frac{2}{r}[\sum_{j=1}^{m} P(Y_j > t) - \frac{1}{r-1}\sum_{i<j} P(Y_i > t, Y_j > t)] \tag{5.33}$$

where r is the greatest integer less than or equal to

$$\sum_{i \neq j} P(Y_i > t, Y_j > t) / \sum_{j=1}^{m} P(Y_j > t); \tag{5.34}$$

and (iii) Kounias (1968),

$$P(\max_{1 \leq j \leq m} Y_j \leq t) \leq 1 - \{ \sum_{j \in L} P(Y_j > t) - \sum_{\substack{i<j \\ i,j \in L}} P(Y_i > t, Y_j > t) \tag{5.35}$$

where L is any subset of {1, 2, ..., m} with two or more elements.

A lower bound, proposed by Hunter (1976), is

$$P(\max_{1 \leq j \leq m} Y_j \leq t) \geq 1 - \sum_{j=1}^{m} P(Y_j > t) - \sum_{(i,j) \in T} P(Y_i > t, Y_j > t) \tag{5.36}$$

where T is any connected set of m - 1 pairs (i,j) such that either (.,k) or (k,.) is in the set for each k = 1, ..., m.

The primary difficulty in evaluating these bounds is that the subnetwork's paths must be explicitly enumerated in order to compute the $P(Y_i > t, Y_j > t)$. Should this computational difficulty be overcome, however, the bounds could be incorporated into the Subnetwork Analysis procedure as per the discussion in subsection 5.3.3.

6. CONCLUDING REMARKS

This report had as its goal the improvement and implementation of a new project scheduling procedure currently being developed at the Institute of Statistics, Texas A&M University. The project scheduling procedure has been improved by significantly extending the very critical Subnetwork Analysis procedure. In particular, a suitable sampling procedure and estimator for bounds on the subnetwork's duration distribution, $\bar{F}(t)$, has been developed and incorporated. In addition, a procedure for extrapolating upper and lower bounds on $\bar{F}(t)$ to obtain an estimate of $\bar{F}(t)$ has also been determined and implemented.

A computer system implementing the project scheduling procedure (including the improvements in Subnetwork Analysis) has been prepared and is documented in Baker and Sielken (1978).

In addition, some possible alternatives to the current Subnetwork Analysis procedure have been suggested. These alternatives are interesting topics for future research.

The authors wish to acknowledge their gratitude to the Office of Naval Research for the support of this research under contracts N00014-68-A-0140 and N00014-76-C-0038. Several present and past members of the Institute of Statistics at Texas A&M University have also contributed to the development of this new project scheduling procedure and its computer implementation: E. Arseven (Lederle Laboratories), P. P. Biemer, C. S. Dunn, N. E. Fisher (Compucon Inc.), L. J. Ringer, and R. K. Spoeri (Bureau of the Census). The authors also want to particularly acknowledge the considerable contributions of H. O. Hartley to Statistical PERT.

REFERENCES

Baker, T. C. Jr. and Sielken, R. L. Jr. (1978). A user's guide to the computer implementation of the new project scheduling procedure: STATISTICAL PERT. THEMIS Project Technical Report No. 57, Institute of Statistics, Texas A&M University, College Station, Texas.

Burt, J. M. Jr. and Garman, M. B. (1971). Conditional Monte Carlo: A simulation technique for stochastic network analysis. Management Science 18, 207-217.

Chung, K. L. and Erdos, P. (1952). On the application of the Borel-Cantelli lemma. Transactions of the American Mathematical Society 72, 179-186.

Clingen, C. T. (1964). A modification of Fulkerson's algorithm for expected duration of a PERT project when activities have continuous d.f. Operations Research 12, 629-632.

Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. Annals of Mathematical Statistics 17, 164-177.

Dawson, D. A. and Sankoff, D. (1967). An equality for probabilities. Proceedings of the American Mathematical Society 18, 504-507.

Devroye, L. P. (1978). Inequalities for the completion times of stochastic PERT networks. Private communication.

Dunn, C. S. and Sielken, R. L. Jr. (1977). Statistical PERT: An improved project scheduling algorithm. THEMIS Project Technical Report No. 55, Institute of Statistics, Texas A&M University, College Station, Texas.

Elmaghraby, S. E. (1967). On the expected duration of PERT type networks. Management Science 13, 299-306.

Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of random variables, with applications. Annals of Mathematical Statistics 38, 1466-1474.

Fulkerson, D. R. (1962). Expected critical path lengths in PERT networks. Operations Research 10, 808-817.

Gaver, D. P. Jr. and Burt, J. M. Jr. (1968). Simple stochastic networks: Some problems and procedures. Management Science Research Report No. 192, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburg, Pennsylvania.

Hartley, H. O. and Wortham, W. A. (1966). A statistical theory for PERT critical path analysis. Management Science 12: 10, 469-481.

Hunter, D. (1976). An upper bound for the probability of a union. Journal of Applied Probability 13, 597-603.

Kleindorfer, G. B. (1971). Bounding distributions for a stochastic acyclic network. Operations Research 19, 1586-1601.

Kounias, E. (1968). Bounds for the probability of a union, with applications. Annals of Mathematical Statistics 39, 2154-2158.

Malcolm, D. G., Roseboom, J. H. and Fazar, W. (1959). Applications of a technique for R&D program evaluation. Operations Research 7, 646-669.

Martin, J. J. (1965). Distribution of time through a directed, acyclic network. Operations Research 13, 46-66.

Moder, J. and Phillips, C. (1974). Project Management with CPM and PERT. Second edition. Van Nostrand-Reinhold Company, New York, New York.

Peizer, D. B. and Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common related tail probabilities, I. Journal of the American Statistical Association 63, 1416-1456.

Ringer, L. J. (1969). Numerical operators for statistical PERT critical path analysis. Management Science 16:2, B-136-B-143.

Robillard, P. and Trahan, M. (1977). The completion time of PERT networks. Operations Research 25, 15-29.

Sielken, R. L. Jr. and Fisher, N. E. (1976). Statistical PERT: Decomposing a project network. THEMIS Project Technical Report No. 50, Institute of Statistics, Texas A&M University, College Station, Texas.

Sielken, R. L. Jr. and Hartley, H. O. (1977). A new statistical approach to project scheduling. THEMIS Project Technical Report No. 56, Institute of Statistics, Texas A&M University, College Station, Texas.

Sielken, R. L. Jr., Hartley, H. O. and Spoeri, R. K. (1976). Statistical PERT: An improved subnetwork analysis procedure. THEMIS Project Technical Report No. 51, Institute of Statistics, Texas A&M University, College Station, Texas.

Sielken, R. L. Jr., Ringer, L. J., Hartley, H. O. and Arseven, E. (1974). Statistical critical path analysis in acyclic stochastic networks: Statistical PERT. THEMIS Project Technical Report No. 48, Institute of Statistics, Texas A&M University, College Station, Texas.

Van Slyke, R. M. (1963). Monte Carlo methods and the PERT problem. Operations Research 11, 839-860.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Texas A&M University

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
Unclassified

3. REPORT TITLE
STATISTICAL PERT: IMPROVEMENTS IN THE DETERMINATION OF THE PROJECT COMPLETION TIME DISTRIBUTION

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*
Thomas C. Baker, Jr. and Robert L. Sielken Jr.

6. REPORT DATE
August 1978

7a. TOTAL NO. OF PAGES
96

7b. NO. OF REFS
27

8a. CONTRACT OR GRANT NO.
N00014-78-C-0426

b. PROJECT NO.
NR047-179

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Report No. 58

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research

13. ABSTRACT

This report develops improvements to a new project scheduling procedure, Statistical PERT, being developed at the Institute of Statistics, Texas A&M University. The project scheduling algorithm is a five step iterative procedure capable of determining a minimum cost project schedule when the activities making up the project have durations which are random variables. The cost of an activity is assumed to be a convex piecewise linear function of the activity's mean duration. The problem is to determine the activity mean durations which both minimize the total project cost and insure that the mean (or some specified percentile) of the corresponding project completion time distribution is less than or equal to a specified project deadline. The entire distribution of the project's completion time under the minimum cost schedule is a valuable by-product.

A critical step, Subnetwork Analysis, in the proposed procedure is improved and extended. Subnetwork Analysis determines an estimate of the duration distribution, $F(t)$, for each subnetwork identified in the previous steps. This estimate is extended to include an extrapolation of upper and lower bounds on $F(t)$. This report also develops a new sampling procedure which results in improved estimators for the bounds on $F(t)$.

ATTACHMENT III

DD FORM 1 NOV 68 1473 (PAGE 1)

S/N 0101-807-6801

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Statistical PERT						
Project Scheduling						
Minimum Cost Scheduling						
Random Activity Durations						
Project Completion Time Distribution						
Network Duration Distribution						
Bounding Network Duration Distributions						

ATTACHMENT III (Continued)

BASIC DISTRIBUTION LIST FOR
UNCLASSIFIED TECHNICAL REPORTS

	Copies
Operations Research Office of Naval Research (Code 434) Arlington, Virginia 22217	3
Director, Naval Research Laboratory Attn: Library, Code 2029 (ONRL) Washington, D.C. 20390	2
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12
Defense Logistics Studies Information Exchange Army Logistics Management Center Fort Lee, Virginia 23801	1
Technical Information Center Naval Research Laboratory Washington, D.C. 20390	6
Office of Naval Research New York Area Office 207 West 24th Street New York, New York 10011 Attn: Dr. J. Laderman	1
Director, Office of Naval Research Branch Office 1030 East Green Street Pasadena, California 91101 Attn: Dr. A.R. Laufer	1
Yale University Department of Statistics Box 2179 - Yale Station New Haven, CT 06520 Attn: Prof. I.R. Savage	1
Mr. William Dejka Code 360 Naval Electronics Laboratory Center San Diego, California 92132	1
Professor Lucien A. Schmit, Jr. Room 6731, Boelter Hall School of Engrng. & Applied Sci. University of California Los Angeles, California 90024	1

	Copies
Director Office of Naval Research Branch Office 495 Summer Street Boston, Massachusetts 02210 Attn: Dr. A.L. Powell	1
Prof. Emanuel Parzen Institute of Statistics Texas A&M University College Station, TX 77843	1
Office of Naval Research San Francisco Area Office 760 Market St. - Room 447 San Francisco, California 94103	1
Technical Library Naval Ordnance Station Indian Head, Maryland 20640	1
Bureau of Naval Personnel Navy Department Technical Library Washington, D.C. 20370 STOP 82	2
Library, Code 0212 Naval Postgraduate School Monterey, California 93940	1
Naval Ordnance Station Louisville, Kentucky 40214	1
Library Naval Electronics Laboratory Center San Diego, California 92152	1
Naval Ship Engineering Center Philadelphia Division Technical Library Philadelphia, Pennsylvania 19112	1
Dr. A.L. Slafkosky Scientific Advisor Commandant Marine Corps (Code AX) Washington, D.C. 20380	1

	Copies
Purdue University Graduate School of Ind. Admin. Lafayette, Indiana 47907 Attn: Prof. Andrew Whinston	1
Department of Economics Tisch Hall - 5th Floor Washington Square, NY Univ. New York City, NY 10003 Attn: Prof. O. Morgenstern	1
Dept. of Statistics Syntex Research 3401 Hillview Palo Alto, CA 94304 Attn: Prof. Stuart Bessler	1
University of California Department of Engineering Los Angeles, California 90024 Attn: Prof. R.R. O'Neill	1
Maritime Transportation Research Board National Academy of Sciences 2101 Constitution Avenue Washington, D.C. 20418 Attn: RADM J.B. Oren USCG (RET)	1
Stanford University Department of Operations Research Stanford, California 94305 Attn: Prof. A.F. Veinott, Jr.	1
Texas A&M Foundation College Station, Texas 77843 Attn: Prof. H.O. Hartley	1
Case Western Reserve University Cleveland, Ohio 44106 Attn: Prof. B.V. Dean	1
University of California Center for Research in Management Science Berkeley, California 94720 Attn: Prof. W.I. Zangwill	1
U.S. Naval Postgraduate School Department of Operations Research and Economics Monterey, California 93940 Attn: Prof. C.R. Jones	1

	Copies
Dr. Del Cilmer Naval Weapons Center (Code 607) China Lake, California 93555	1
Dr. Paul Murrill Project Themis Department of Chemical Engineering Louisiana State University Baton Rouge, Louisiana 70803	1
Director Army Materials & Mechanics Research Center Attn: Mr. J. Bluhm Watertown, Massachusetts 02172	1
Commanding Officer U.S. Army Ballistic Research Laboratories Attn: Dr. John H. Giese Aberdeen Proving Ground, Maryland 21005	1
University of Iowa Department of Mechanics and Hydraulics Iowa City, Iowa 52240 Attn: Prof. W.F. Ames	1
Office of Naval Research Resident Representative 1105 Guadalupe Lowich Building Austin, Texas 78701 Attn: Mr. Francis M. Lucas	1
Dr. Jerome Bracken Institute of Defense Analyses 400 Army-Navy Drive Arlington, Virginia 22202	1
Professor Richard L. Fox School of Engineering Case Western Reserve University Cleveland, Ohio 44106	1
	1

	Copies
	1
	1
University of Chicago Statistical Research Center Chicago, Illinois 60637 Attn: Prof. W. Kruskal	1
Stanford University Department of Statistics Stanford, California 94305 Attn: Prof. G.J. Lieberman	1
	1
Princeton University Department of Mathematics Princeton, New Jersey 08540 Attn: Prof. J.W. Tukey	1
Stanford University Department of Statistics Stanford, California 94305 Attn: Prof. T.W. Anderson	1
University of California Department of Statistics Berkeley, California 94720 Attn: Prof. P.J. Bickel	1
University of Washington Department of Mathematics Seattle, Washington 98105 Attn: Prof. Z.W. Birnbaum	1
Cornell University Department of Mathematics White Hall Ithaca, New York 14850 Attn: Prof. J. Kiefer	1
Harvard University Department of Statistics Cambridge, Massachusetts 02139 Attn: Prof. W.G. Cochran	1
Columbia University Department of Industrial Engineering New York, New York 10027 Attn: Prof. C. Derman	1
New York University Institute of Mathematical Sciences New York, New York 10453 Attn: Prof. W.M. Hirsch	1
University of North Carolina Statistics Department Chapel Hill, North Carolina 27515 Attn: Prof. W.L. Smith and Prof. M.R. Leadbetter	1
Purdue University Division of Mathematical Sciences Lafayette, Indiana 47079 Attn: Prof. H. Rubin	1
University of California, San Diego Department of Mathematics P.O. Box 109 La Jolla, California 92038 Attn: Prof. M. Rosenblatt	1
Florida State University Department of Statistics Tallahassee, Florida 32306 Attn: Prof. R.A. Bradley	1
New York University Department of Industrial Engineering & Operations Research Bronx, New York 10453 Attn: Prof. J.H.K. Kao	1
University of Wisconsin Department of Statistics Madison, Wisconsin 53706 Attn: Prof. G.E.P. Box	1

	Copies
Logistics Research Project The George Washington University 707 - 22nd Street, N.W. Washington, D. C. 20037 Attn: Dr. W. H. Marlow	1
International Business Machines Corporation P.O. Box 218, Lamb Estate Yorktown Heights, New York 10598 Attn: Dr. Alan Hoffman	1
University of California Management Sciences Research Project Los Angeles, California 90024 Attn: Dr. J.R. Jackson	1
Harvard University Department of Economics Cambridge, Massachusetts 02138 Attn: Prof. K.J. Arrow	1
Cowles Commission for Research in Economics Yale University New Haven, Connecticut 06520 Attn: Prof. Martin Shubik	1
Carnegie-Mellon University Graduate School of Industrial Administration Pittsburgh, Pennsylvania 15213 Attn: Prof. G. Thompson	1
University of California Department of Economics Berkeley, California 94720 Attn: Prof. R. Radner	1
University of California Operations Research Center Institute of Engineering Research Berkeley, California 94720 Attn: Prof. D. Gale	1
University of California Graduate School of Business Administration Los Angeles, California 90024 Attn: Prof. J. Marschak	1
The University of Michigan Department of Mathematics, W.E. Ann Arbor, Michigan 48104 Attn: Prof. R.M. Thrall	1
Princeton University Department of Mathematics Princeton, New Jersey 08540 Attn: Prof. A.W. Tucker	1
Case Western Reserve University Systems Research Center Cleveland, Ohio 44106 Attn: Prof. M. Mesarovic	1
University of Texas Department of Mathematics Austin, Texas 78712 Attn: Dr. A. Charnes	1
Stanford University Department of Operations Research Stanford, California 94305 Attn: Dr. D.L. Iglehart	1
University of Delaware Department of Mathematics Newark, Delaware 19711 Attn: Prof. R. Remage, Jr.	1
Stanford University Department of Operations Research Stanford, California 94305 Attn: Prof. F.S. Hillier	1
Dr. Claude-Alain Burdet Asst. Prof. Industrial Admin. Carnegie-Mellon University Pittsburgh, Pennsylvania 15213	1
Stanford University Department of Operations Research Stanford, California 94305 Attn: Prof. G. B. Dantzig	1
Chief of Naval Research (Code 436) Department of the Navy Arlington, Va. 22217	1
Science Librarian Kresge Library Oakland University Rochester, Michigan 48063	1

	Copies
University of Connecticut Department of Statistics Storrs, Connecticut 06268 Attn: Prof. H.O. Posten	1
ARCON Corporation Lakeside Office Park North Avenue at Route 128 Wakefield, Massachusetts 01880 Attn: Dr. A. Albert	1
Stanford University Department of Statistics Stanford, California 94305 Attn: Prof. H. Chernoff	1
Yale University Department of Statistics New Haven, Connecticut 06520	1
Rutgers-The State University Statistics Center New Brunswick, New Jersey 08903 Attn: Prof. H.F. Dodge	1
Yale University Department of Statistics New Haven, Connecticut 06520 Attn: Prof. F. J. Anscombe	1
Purdue University Division of Mathematical Sciences Lafayette, Indiana 47907 Attn: Prof. S.S. Gupta	1
Cornell University Department of Industrial Engineering Ithaca, New York 19850 Attn: Prof. R.E. Bechhofer	1
Mrs. Barbara Eaudi Univ. Program Coordinator, B.E. NASA Johnson Space Center Houston, TX 77058	1
Professor Geoffrey S. Watson, Chairman Department of Statistics Princeton University Fine Hall Princeton, N. J. 08540	1
Stanford University Department of Mathematics Stanford, California 94305 Attn: Prof. S. Karlin	1
University of Sheffield Department of Probability and Statistics Sheffield 10, ENGLAND Attn: Prof. J. Gani	1
University of California Operations Research Center Institute of Engineering Research Berkeley, California 94720 Attn: Prof. R.E. Barlow	1
Stanford University Department of Statistics Stanford, California 94305 Attn: Prof. H. Solomon	1
Applied Mathematics Laboratory Naval Ships Research Development Center Washington, D.C. 20007	1
Systems Analysis Division Room BE760, Pentagon Washington, D.C. 20350 Attn: Mr. A.S. Rhodes, Op-964	1
Department of Statistics University of North Carolina Chapel Hill, North Carolina 27515 Attn: Prof. M.R Leadbetter	1
Southern Methodist University Department of Statistics Dallas, Texas 75222 Attn: Prof. D.B. Owen	1
Israel Institute of Technology Technion Haifa, ISRAEL Attn: Prof. P. Naor	

Office of Gifts & Exchanges Library Texas A&M University College Station, Texas 77843	Copies 2
Archives Texas A&M University College Station, TX 77843	 1